RecPad 2014
20^{th} Portuguese Conference on Pattern Recognition

Universidade da Beira Interior, Covilhã, Portugal

October 31^{st}, 2014

UNIVERSIDADE DA BEIRA INTERIOR
Covilhã | Portugal

Contents

Committees

Organizing Committee

- Luís A. Alexandre ,General Chair (UBI)
- Hugo Proença (UBI)
- Paulo Fazendeiro (UBI)

Technical Committee

- Manya Afonso (IST)
- Hélder Araújo (UC)
- Jorge Barbosa (FEUP)
- Alexandre Bernardino (IST)
- Hans du Buf (UAlg)
- Aurélio Campilho (FEUP)
- Jaime Cardoso (FEUP)
- Paulo Carvalho (UC)
- Joaquim Pinto da Costa (FCUP)
- José Bioucas Dias (IST)
- Mário A. T. Figueiredo (IST)
- Ana Fred (IST)
- Noel Lopes (IPG)

- Ana Maria Mendonça (FEUP)
- Pedro Pina (IST)
- António Pinheiro (UBI)
- Armando Pinho (UA)
- Pedro Quelhas (INEB)
- Bernardete Ribeiro (UC)
- João Rodrigues (UAlg)
- Paulo Salgado (UTAD)
- João Sanches (IST)
- Beatriz Sousa Santos (UA)
- Jorge Santos (ISEP)
- Catarina Silva (IPLeiria)
- Luís Silva (UA)
- José Silva (Academia Militar)
- Ricardo Sousa (INEB)
- João Tavares (FEUP)
- Luíís F. Teixeira (FEUP)
- Ana Maria Tomè (UA)
- Verónica Vasconcelos (ISEC)
- Andrzej Wichert (IST)

Foreword

The 20th edition of the Portuguese Conference on Pattern Recognition, RECPAD 2014, is held at the Universidade da Beira Interior, Covilhã, Portugal on October the 31st, 2014.

It is a great honor for UBI and for the members of the Organizing Committee to have this opportunity to organize this conference on the year of its 20th anniversary.

We received 41 submissions of which 39 were accepted. All submissions were double blind and were reviewed by two members of the Technical Committee. The conference will have a Best Poster Award and also the ceremony of the APRP Master Thesis Award.

We will have an invited lecture by Prof. Francesc Moreno-Noguer, who is an Associate Researcher at the Institut de Robòtica i Informàtica Industrial (CSIC-UPC), Barcelona, Spain. Prof. Moreno-Noguer will present a talk on Monocular 3D detection of rigid and non-rigid shapes.

We are very happy to have the support of the following sponsors, which made possible for us to keep the registration fees low: EyeSee Solutions (Platinum sponsor), Indra and InovaPrime (both Gold sponsors).

On behalf of the organising committee we would like to thank all the people involved in this event, namely, the members of the Technical Committee, the Portuguese Association for Pattern Recognition, APRP, specially its president, Prof. Jaime S. Cardoso and the staff at the Royal Veiga Factory Museum at UBI, which will held the conference. The help of Mrs. Dulce Serralheiro is also gratefully acknowledged.

Above all, the conference is made by the participants and our final words are to thank you all for submitting your work and for coming to UBI on the celebration of the 20 years of RecPad. We hope you enjoy this year's edition of RecPad.

The organizing committee.

Sponsors

Platinum sponsor:

Gold sponsors:

Papers Links

- *Noninvasive Blood Pressure and the Second Heart Sound Analysis*; Ana Castro, Sandra S. Mattos and Miguel Coimbra.
- *Periodic background pattern detection and removal for cell tracking*; Tiago Esteves, Ângela Carvalho, Fernando Jorge Monteiro and Pedro Quelhas
- *Handwritten Signature Matching using GPUMLib*; João Gonçalves, Noel Lopes and Bernardete Ribeiro.
- *Border Abruptness Assessment Methodology for Skin Lesion Image Analysis*; Maria João M. Vasconcelos, Luís Rosado and Márcia Ferreira.
- *Classification of Lung Function on a Smartphone App*; João Teixeira, Luís Teixeira, João Fonseca and Tiago Jacinto.
- *Blind Inpainting after Log Transformation using Total Variation*; Manya V. Afonso and João M. R. Sanches.
- *Quis-Campi: Extending In The Wild Biometric Recognition to Surveillance Environments*; João Neves and Gil Santos.
- *Partial Face Recognition Under Occlusion Using Universal Background Models*; João C. Monteiro and Jaime S. Cardoso.
- *Liveness detection methods in fingerprint recognition*; Ana F. Sequeira, Ana Rute Louro and Jaime S. Cardoso.
- *Deteção de Veículos e Edifícios*; Jorge Leitão, José Bioucas-Dias, José S. Silva and Miguel Gonçalves
- *Discriminação da densidade vegetacional em ambiente militar*; Miguel Gonçalves, José Bioucas-Dias, José S. Silva and Jorge Leitão.
- *Metadata Extraction from Ancient Documents*; Henrique Mata, João R. Caldas Pinto and Maria C. S. Nunes.

- *Attenuation of speckle noise in OCT images*; Fabiana Rodrigues, José Silvestre Silva and Jaime Santos.
- *Safety Level opposite Leakage in underground fuel tank based on the Model of Pattern Recognition*; Erico Gaspar Lisboa.
- *Optical Character Recognition of Guinea and Bissau boundaries documentation*; Henrique Mata, João R. Caldas Pinto and Maria C. S. Nunes.
- *Toward Building an Automatic Video Surveillance System*; Samaneh Khoshrou, Jaime S. Cardoso and Luís F. Teixeira.
- *A rate-distortion study on microarray image compression*; Luís M. O. Matos, António J. R. Neves and Armando J. Pinho.
- *Dynamic Sub-Sampling Line Scan Visual Inertial Odometry with Application to Mobile Robots*; Ricardo Silva, Vitor Santos and Ricardo Pascoal.
- *RetinaCAD - Retinal Computer-Aided Diagnosis System*; Behdad Dashtbozorg, Ana Maria Mendonça, Susana Penas, Jorge Polónia and Aurélio Campilho.
- *Using Support Vector Machine model for fault detection along a water canal*; José Duarte, Luís Rato and Manuel Rijo.
- *Is it Possible to estimate the Embedding Dimension using FNN?*; Jorge Henrique Santos Oliveira, Miguel Coimbra.
- *Protein quantification of fluorescence images in heterogeneous cell population*; Cláudia Gomes, Raquel Seruca and João Sanches.
- *Improving Classification Accuracy of Deep Neural Networks by Transferring Features from a Different Distribution*; Chetak Kandaswamy, Luís M. Silva and Jaime S. Cardoso.
- *Transfer Learning: Current Status, Trends and Challenges*; Ricardo Sousa, Luís M. Silva, Luís A. Alexandre, Jorge Santos and Joaquim Marques de Sá.
- *CorSiL: A Novel Dataset for Portuguese Sign Language and Expressiveness Recognition*; Pedro M. Ferreira, Inês V. Rodrigues, Ana Rio, Ricardo Sousa, Eduardo M. Pereira and A. Rebelo.
- *Efficient Supervised Relevance Criteria Tailored to Discrete Features*; Artur J. Ferreira and Mário A. T. Figueiredo.

- *A comparison between Shannon's, Renyi's, and Tsallis Mutual Information for Feature Selection*; Artur J. Ferreira and Mário A. T. Figueiredo.
- *On-Line Evaluation of Open-Ended Object Recognition System*; Hamidreza Kasaei, Miguel Oliveira, Gi Hyun Lim, Luís Seabra Lopes and A. M. Tomé.
- *Segmentation and Tracking of the Mitral Heart Valve in 3D Echocardiography*; Malik Saad Sultan and Miguel T. Coimbra.
- *Learning with Drift in Twitter*; Joana Costa, Catarina Silva, Mário Antunes and Bernardete Ribeiro.
- *Detecting Unknown Pattern Repeats: First Results Using a Probabilistic Approach*; Raquel Sebastião and Armando J. Pinho.
- *Non-rigid Registration of Point Clouds Regularized by NURBS*; Hooshiar Zolfagharnasab and Hélder P. Oliveira.
- *Multi-Target Tracking Based in Meanshift and Particle Filters*; Pedro Silva, Ricardo Ferreira and José Gaspar.
- *Large-scale inversions between human reference assemblies*; Diogo Pratas, Raquel M. Silva and Armando J. Pinho.
- *Towards personalized medicine: ebola virus absent words in the human genome*; Raquel M. Silva, Luisa Castro, Diogo Pratas and Armando J. Pinho.
- *A Cellular Automaton for Enzyme Kinetics*; João Silva, Nelson Fernandes, Paulo Fazendeiro and António Mendonça..
- *Automatic Segmentation of Corneal Nerves from In Vivo Confocal Microscopy images using Chan-Vese Model and Tortuosity Evaluation*; David Saraiva, Hugo Marcos, Verónica Vasconcelos and António Luís Marques.
- *Single-Frame Image Denoising and Inpainting using Gaussian Mixture Models and Minimum Mean Squared Error Estimation*; Afonso Teodoro, Mariana Almeida, Mário A. T. Figueiredo.
- *A 3D Keypoint Detector based on Biologically Motivated Bottom-Up Saliency Map*; Sílvio Filipe and Luís A. Alexandre.

Papers

Noninvasive Blood Pressure and the Second Heart Sound Analysis

Ana Castro
ana.castro@dcc.fc.up.pt

Sandra S. Mattos
ssmattos@cardiol.br

Miguel T. Coimbra
mcoimbra@dcc.fc.up.pt

Faculdade de Engenharia da Universidade do Porto
Instituto de Telecomunicações
Portugal

Unidade de Cardiologia e Medicina Fetal
Real Hospital Português
Recife, Brasil

Faculdade de Ciências da Universidade do Porto
Instituto de Telecomunicações
Portugal

Abstract

Auscultations acquired from children patients (27 patients, 10.2±3.9 years, 35.7±20.8 kg, 132.3±25.5 cm), were automatically segmented into their components first (S1) and second (S2) heart sounds. A set of time, frequency, and wavelet based features, were extracted from the S2, and analyzed in relation to the noninvasive cuff-based measures of blood pressure (mean blood pressure of 78±8.8 mmHg). A multivariate regression analysis was performed to determine which features better related to the blood pressure measurements. The best results, in the leave-one-out evaluation, were obtained using the frequency features set (MAE 6.08 mmHg, MAPE 7.85%, ME 0.31 mmHg).

1 Introduction

Standard methods of blood pressure (BP) monitoring are uncomfortable and inadequate for long term evaluation, leading to the search of new techniques applying different physiological properties of the cardiovascular system, such as the Pulse Transit Time (PTT) [9], vascular transit time (VTT) [6], and the heart sound [3, 8].

In this study, time-frequency features extracted from the second heart sound (S2) were analyzed in relation to the noninvasive peripheral BP measures.

2 MATERIAL AND METHODS

Data collected with a Littmann® 3200 electronic stethoscope (4 kHz), and the DigiScope Collector, from 27 patients, 10 female. Data were anonymized and shipped to Portugal with the approval of the RHP and University of Porto Ethics Commitees. Noninvasive systemic BP was measured in the arm and sitting position (Table 1), and Mean Blood Pressure (MBP) was approximated according to Equation 1 [1]. PCGs collected over the pulmonary auscultation spot were segmented as described in [5]. Total auscultation time was of 307.5 s (432 detected S2s).

$$MBP \approx \frac{2DBP + SBP}{3} \tag{1}$$

.

2.1 Heart Sound Features Extraction and Blood Pressure

Detected S2s were analyzed for each patient, and time and frequency based features extracted. Temporal contour envelopes were extracted, based on the local maximums and minimums of the signal, and interpolated using a cubic Hermitian spline. For each positive envelope, the two largest peaks are detected (estimated A2 and P2 components). The lag between peaks was also extracted. Similarly, the energy of the signal was obtained (Equation 2), and the peaks of the energy envelope extracted as before (A2, P2, and lag). The difference between positive and negative contour envelopes was used to extract periods of sound and silence (10% threshold), storing S2 duration and the zero-crossing rate.

$$E = \frac{1}{win} \sum_{i=1}^{win} PCG(i)^2 \tag{2}$$

where *win* is the window length.

Shannon entropy was also extracted from each S2 (Equation 3).

$$Entropy = \frac{1}{n} \sum_{i=1}^{n} p_i log(p_i) \tag{3}$$

for a set of events with PDF $\{p_i, i = 1, ..., n\}$.

Following time-based analysis, frequency analysis of each S2 segment was performed extracting spectral roll-off (85%), centroid, and peak frequency [2]. Short-time spectral analysis was performed on the S2, subdivided into smaller windows, and the spectrum estimated (Hamming window). The spectral features previously referred were then extracted from the maximum energy window. Order 6 Daubechies wavelet transform was also obtained, and energy distribution over the approximation and details was extracted and incorporated in the features matrix [5].

Features were extracted from each heart sound, for each patient, and median values of all detected S2s were assigned to each patient. Features were used in a multivariate linear regression model of the MBP (Equation 4), and adjusted for the total data using the three different features' sets (least square error). For each features' set a stepwise feature selection was performed, and a linear multivariate model adjusted and evaluated using a leave-one-out approach (reduced number of data points available).

$$\widehat{MBP}_i = a_0 + a_1 x_1(i) + ... + a_l x_l(i) + e(i) \tag{4}$$

where a_i is the estimated weight of attribute x_i on the estimation $\widehat{MBP}_i$.

2.2 Evaluation Criteria

According to the British Hypertension Society, requirements for Grade A BP monitors are that measurements within a 5 mmHg range includes at least 60% of data, measurements within 10 mmHg comply 85% of data, and measurements within 15 mmHg range include 95% of the data [7]. Cumulative percentages, Mean absolute error (MAE, Equation 5), mean absolute percentage error (MAPE, Equation 6), mean error (ME, Equation 7), and error standard-deviation were calculated. According to the US Association for the Advancement of Medical Instrumentation standard, a BP device must display MAE inferior to 5 mmHg, and an error standard-deviation inferior to 8 mmHg [9]. For each set linear correlation coefficient was obtained, and a Bland-Altman analysis performed [4].

$$MAE = \frac{1}{m} \sum_{i=1}^{m} \left| MBP_i - \widehat{MBP}_i \right| \tag{5}$$

$$MAPE(\%) = \frac{1}{m} \sum_{i=1}^{m} \frac{\left| MBP_i - \widehat{MBP}_i \right|}{MBP_i} \tag{6}$$

$$ME = \frac{1}{m} \sum_{i=1}^{m} MBP_i - \widehat{MBP}_i \tag{7}$$

where m is the number of patients.

3 RESULTS AND DISCUSSION

Table 2 presents the results obtained for each model regarding correlation, and estimation errors, and Figure 1 presents the Bland-Altman plots for the error analysis of the frequency features set model, which exhibited lower error. The best results in the cross-validation were obtained for the model using the S2 spectral features, which relation has also been

Table 1: Sample mean, standard-deviation, and range values of the demographic data, and blood pressure measurements.

	Age (years)	Weight (kg)	Height (cm)	SBP (mmHg)	DBP (mmHg)	MBP (mmHg)
Mean	10.2	35.7	132.3	104.6	64.7	78.0
Standard-Deviation	3.9	20.8	25.5	12.9	8.4	8.8
Range (min-max)	3-18	8-97	69-183	90-145	50-90	63.3-103.3

Table 2: Model results of estimated blood pressure, using S2 characteristics, in relation to measured blood pressure: correlation coefficient (ρ), mean absolute error (MAE), mean absolute percentage error (MAPE), mean error (ME), error standard-deviation (SD), and percentage estimates within error bands of 5, 10 and 15 mmHg

						% Estimates Within Error Band		
Features	ρ	MAE (<5)	MAPE (%)	ME	Error SD (<8)	5 mmHg (>60%)	10 mmHg (>85%)	15 mmHg (>95%)
Total Data Linear Model								
Total	0.82	4.11	5.28	0.00	5.03	62.96	100.00	100.00
Time	0.40	5.91	7.48	0.00	8.05	48.15	85.19	92.59
Frequency	0.60	5.03	6.54	0.00	7.03	66.67	85.19	96.30
Wavelet	0.36	5.98	7.62	0.00	8.20	55.56	81.48	96.30
Leave-One-Out Adjusted Linear Model								
Total	0.22	7.48	9.69	-0.12	9.98	44.44	70.37	88.89
Time	-0.37	9.18	11.85	-0.88	12.66	37.04	70.37	77.78
Frequency	0.39	6.08	7.85	0.31	8.29	48.15	81.48	88.89
Wavelet	0.23	6.22	7.92	-0.22	8.81	48.15	81.48	92.59

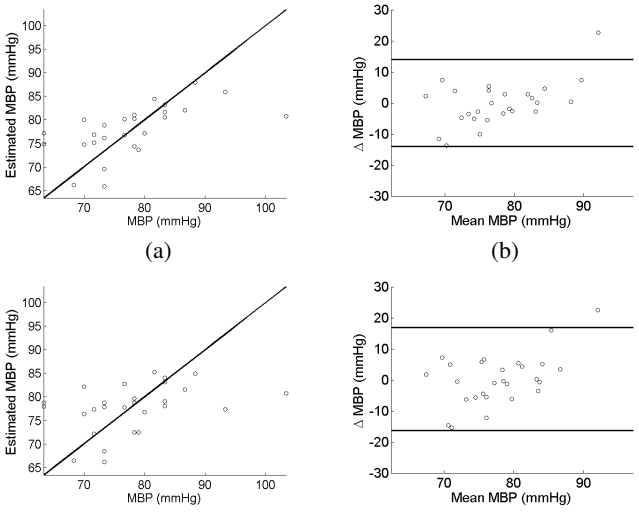


Figure 1: Estimated Mean Blood Pressure (MBP): (a,c) measured MBP versus estimated; (b,d) Bland-Altman plot. (a,b) results for the linear model using all frequency features, and total data; (c,d) results for the leave-one-out adjusted model of frequency features.

reported in the literature [2, 3]. Figure 1, presenting the Bland-Altman analysis, shows that estimations have a wide variation, with no apparent relation between the mean MBP and the error.

Limitations in our study are the reduced number of patients, use of a children population (not homogenous), the fact that only one measure of BP was obtained for each patient not allowing for individual calibration and repeatability analysis, and the limited range of BPs evaluated. Also, the linear multivariate model may not be adequate, since these features may exhibit non-linear relations to the BP [8]. A MAE of 6.08 mmHg was observed in the leave-one-out cross-validation of the frequency features set, which although is below the required standards for BP monitors, present as encouraging results for this exploratory study. Heart sound characteristics may be linked to normal physiological changes during growth, that should be analyzed in detail.

4 CONCLUSIONS

Children PCGs were segmented into S1 and S2 components, and individual S2s analyzed, demonstrating that S2 characteristics are related to the measured MBP. In this exploratory study the spectral features extracted from the S2s presented the best results, comparable to those presented in the literature.

ACKNOWLEDGMENT

This work was partially funded by the Fundação para a Ciência e Tecnologia (FCT, Portuguese Foundation for Science and Technology) under the reference Heart Safe PTDC/EEI-PRO/2857/2012; and Project I-CITY - ICT for Future Health/Faculdade de Engenharia da Universidade do Porto, NORTE-07-0124-FEDER-000068.

References

[1] *Vander's Human Physiology: The Mechanisms of Body Function*. McGraw-Hill, 2001.

[2] P.J. Arnott, G.W. Pfeiffer, and M.E. Tavel. Spectral analysis of heart sounds: Relationships between some physical characteristics and frequency spectra of first and second heart sounds in normals and hypertensives. *Journal of Biomedical Engineering*, 6(2):121 – 128, 1984. ISSN 0141-5425.

[3] Andreas Bartels and Dietrich Harder. Non-invasive determination of systolic blood pressure by heart sound pattern analysis. *Clinical Physics and Physiological Measurement*, 13(3):249 –256, 1992.

[4] J. M. Bland and D. G. Altman. Statistical methods for assessing agreement between two methods of clinical measurement. *Lancet*, 1:307–310, 1986.

[5] Ana Castro, Tiago T. V. Vinhoza, Sandra S. Mattos, and Miguel T. Coimbra. Heart sound segmentation of pediatric auscultations using wavelet analysis. In *Engineering in Medicine and Biology Society, 2013 Proceedings of the 35th Annual International Conference of the IEEE*, 2013.

[6] V. Chandrasekaran, R. Dantu, S. Jonnada, S. Thiyagaraja, and K.P. Subbu. Cuffless differential blood pressure estimation using smart phones. *Biomedical Engineering, IEEE Transactions on*, 60(4): 1080–1089, 2013. ISSN 0018-9294. doi: 10.1109/TBME.2012.2211078.

[7] E. O'Brien, N. Atkins, G. Stergiou, N. Karpettas, G. Parati, R. Asmar, Y. Imai, J. Wang, T. Mengden, and A. Shennan. European society of hypertension international protocol revision 2010 for the validation of blood pressure measuring devices in adults. *Blood Pressure Monitoring*, 15(1):23–38, 2010.

[8] Robert Smith and Dan Ventura. A general model for continuous non-invasive pulmonary artery pressure estimation. *Computers in Biology and Medicine*, 43(7):904 – 913, 2013. ISSN 0010-4825.

[9] J. Sola, M. Proenca, D. Ferrario, J.-A. Porchet, A. Falhi, O. Grossenbacher, Y. Allemann, S.F. Rimoldi, and C. Sartori. Noninvasive and nonocclusive blood pressure estimation via a chest sensor. *Biomedical Engineering, IEEE Transactions on*, 60(12):3505–3513, 2013.

Periodic background pattern detection and removal for cell tracking

Tiago Esteves[1,3]
dee11017@fe.up.pt
Ângela Carvalho[2,3]
angela.carvalho@ineb.up.pt
Fernando Jorge Monteiro[2,3]
fjmont@ineb.up.pt
Pedro Quelhas[1,3]
quelhas@fe.up.pt

[1] Dpt. de Engenharia Electrotécnica e de Computadores, Faculdade de Engenharia, Universidade do Porto, Rua Dr. Roberto Frias, 4200-465 Porto, Portugal
[2] Dpt. de Engenharia Metalúrgica e de Materiais, Faculdade de Engenharia, Universidade do Porto, Rua Dr. Roberto Frias, 4200-465 Porto, Portugal
[3] INEB - Instituto de Engenharia Biomédica, Rua do Campo Algre, 823, 4150-180 Porto, Portugal

Abstract

The study of cell morphology and cell mobility variation when cells are grown on top of patterned substrates is becoming a very important factor in tissue regeneration.

In this paper we present a novel approach to automatically detect and remove periodic background patterns in brightfield microscopy images. This background removal process is fundamental to perform cell tracking as the periodic background pattern would otherwise lead to erroneous cell detection and association. The detection of the background is performed by searching for the periodic background pattern organization through the analysis of keypoints automatically obtained from images. Using this information we are able to both detect and reconstruct the periodic background and finally remove it from the original images.

We tested the proposed approach on microscopy images with different periodic background patterns. The effectiveness of the method was validated by visual inspection and by the cell tracking results obtained.

1 Introduction

The analysis of cell behaviour when interacting with different micropatterned surfaces has gained increasing interest in the last years [1]. Biologist researchers started producing micropatterned surfaces (figure 1) on biomaterials to study the possibility to modulate cell behaviour only through topography stimulus of biomaterials [1]. Those cell/surface interactions are analyzed in order to access cell metabolic activity, adhesion morphology, proliferation and lineage differentiation. Measurements of cell alignment, elongation and guided mobility on the surface are essential to confirm these interactions. Currently, measurements are performed mainly by visual inspection alone because there is not an automatic alternative.

The required cell mobility and morphology analysis is already performed on other studies where the background is flat [2]. Nevertheless it is known that there is an dependency between segmentation and interferences or changes in the image background like changes or distortions of image intensity or illuminance [3]. In order to facilitate cell detection and tracking for mobility and morphology analysis we propose a new approach to automatically remove the periodic background pattern from the original image. The main steps of the proposed approach are: first we detect the background pattern based on its periodicity, if it exists, we synthesize the full background image; finally we subtract the background from the original image obtaining only information related to the existing cells.

2 Proposed methodology

To detect if any periodic pattern exists in the image under analysis we measure the image entropy. A high entropy value, over a predefined *threshold*, indicates that we have high pixel values variation and we assume that it occurs in case of a periodic patterned image. If a background pattern exists the first step for its detection is to extract keypoints from the image that will allow to infer the pattern periodicity, using the Laplacian of Gaussian filter (LoG) [2]. We apply this approach to the images under analysis obtaining keypoints in positions that are related to both the background pattern and cell's position (figure 2). As we observe in images from figure 2, as expected, the keypoints appear on locations with cells. However, they also appear in the regions of the background pattern

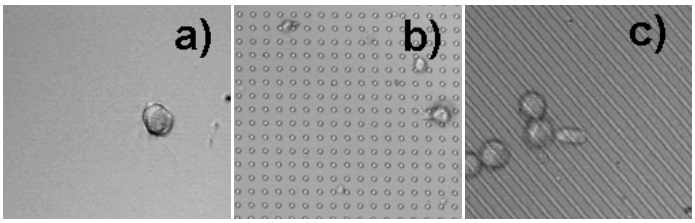


Figure 1: Brightfield images with cells on top of different micropatterned surfaces: a) Flat surface; b) Pillar pattern surface; c) Line pattern surface.

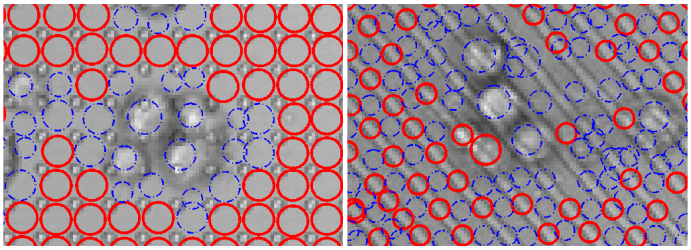


Figure 2: Keypoint extraction from different background pattern brightfield images (line color/shape represent different SIFT clusters).

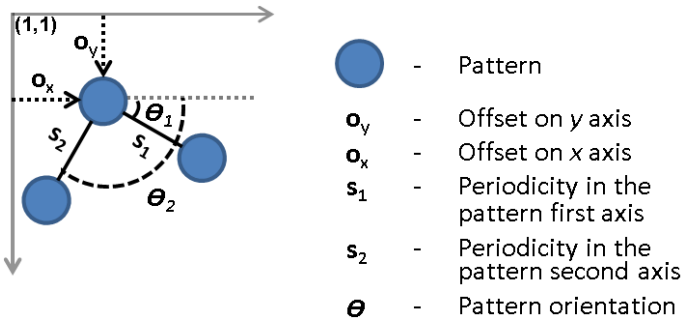


Figure 3: Scheme used for detecting the pattern background periodicity. The background pattern is defined with a specific periodic interval s_1 and s_2, orientations θ_1 and θ_2, and image origin offset defined by (o_x, o_y).

with the same periodicity as the pillars from the background (figure 2 - left) and the same orientation as the lines that compose the background (figure 2 - right).

After obtaining the keypoints we extract for each a SIFT descriptor which we use together with k-means clustering ($k = 4$) to choose the largest cluster of descriptors that corresponds to the keypoints obtained on the pattern repetition (figure 2 - red bold continuous circles). The selected keypoints are used for the pattern periodicity analysis.

2.1 Background pattern periodicity analysis

In order to detect the periodic background pattern we follow the scheme in figure 3. As we consider the periodic pattern to have two orthogonally independent periodicity spacings, we first search only over o_x, o_y, θ_1 and s_1 to find those that best fit our model. Given a set of values for o_x, o_y, θ_1 and s_1 we generate the predicted locations for our periodic pattern along the axis defined by θ_1. Given the selected LoG detection's coordinates in the image, we project each detection's location onto the axis defined by θ_1 along its normal. Each of those projections is then assigned to the nearest

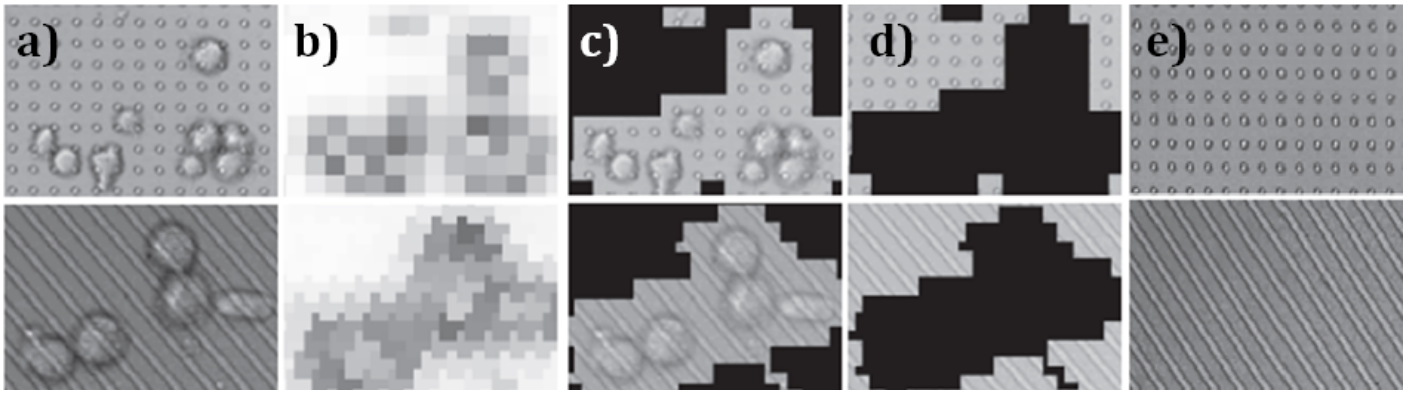


Figure 4: Foreground and background separation: a) Original image; b) similarity map (white high value); c) Foreground regions; d) Background detection; e) Periodic background pattern detection and reconstruction.

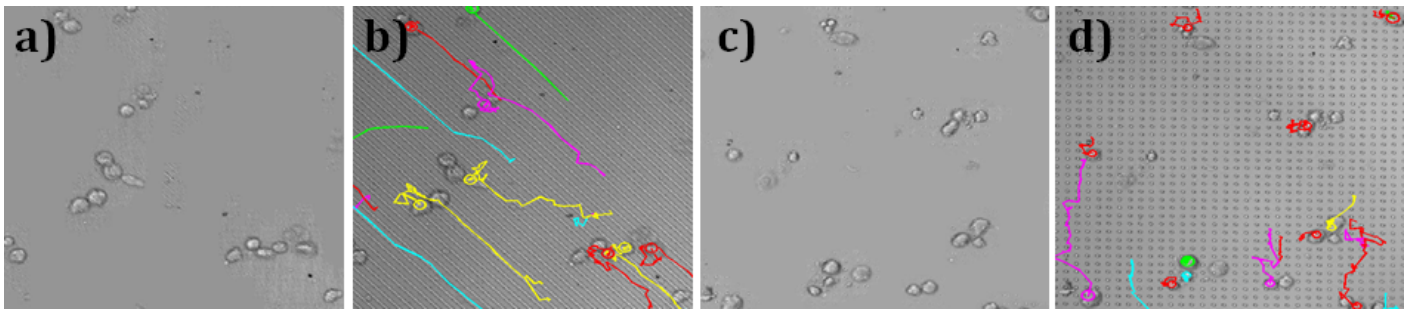


Figure 5: Background removal and cell tracking results: a,c) Result from subtracting the reconstructed background pattern from the original image; b,d) Cell tracks on top of the original image.

periodic pattern location, and the distance to the predicted periodic pattern is computed. As this distance calculation would favor smaller spacing in the periodic pattern it is normalized dividing by s_1. The set of parameters that lead to the lowest average distance are those assumed to represent the periodic pattern of the background in the image. Given o_x, o_y, θ_1 and s_1 we fix o_x, o_y and vary s_2 setting $\theta_2 = \theta_1 \pm 90$, but allowing for a 20 degree tolerance. If in this θ interval we find a second low average distance and the corresponding s_2 value is equal to s_1 value found previously, we assume that the background periodic pattern is pillar type (figure 4 - a, top). We make this assumption because for both θ values (separated by ± 90 degrees) we have the same keypoints periodicity. Otherwise, we assume a line background periodic pattern (figure 4 - a, bottom).

Given the information found we classify each region as being periodic background pattern or foreground using template matching.

2.2 Foreground removal

To detect the foreground regions we measure the similarity of each region with their 8 neighbour regions (according to the pattern periodicity found) using cross correlation.

As we measure the similarity between regions according to the pattern periodicity we will obtain high cross correlation coefficient in presence of background pattern regions, and low values for foreground regions (figure 4 - b). From this analysis we then assume a specific *threshold* that separates the foreground regions (figure 4 - c) from the background pattern (figure 4 - d).

2.3 Background reconstruction

The final step to obtain the entire background pattern image is to reconstruct the foreground regions based on the pattern's periodicity. We define a patch (9×9) and search the image. When the center pixel of the patch is located at a foreground location we synthesize the periodic background pattern and replace it in the image. To synthesize the background we use the similar background patches located according to the parameters (θ_1, θ_2, s_1, s_2, o_x, o_y) previously estimated. We average those similar patches to synthesise a valid background patch for that location and replace it in the background image. If pixels of that patch overlaps with already existing background, those pixels are combined weighting the new pixels with 0.1 and the old pixels with 0.9. Examples of the final results are given in figure 4 - e.

3 Discussion and conclusion

We applied our approach on several brightfield microscopy images with different periodic background patterns and cells on top. For each original image we obtained the periodic background pattern and then we subtracted it from the original image and observed the results (figure 5 - a,c). From the results we were able to visualize that the differences are given mainly due to the cell presence which is an indicator that the background reconstruction is performing well.

Given the removal of the periodic background pattern we use the LoG filter to detect cells and we obtain detections only on locations with cells with no background pattern interference. From the cell detection result the cell tracking is then possible (figure 5 - b,d).

Future work will be done on the analysis of cell morphology and mobility which is now possible based on the development methodology for periodic background pattern removal. The influence of different background pattern (resultant of different micropatterned surfaces) on cell mobility and morphology will be quantified and compared.

Acknowledgements

This work has been financed by FEDER funds through the Programa Operacional Factores de Competitividade - COMPETE and by Portuguese funds through FCT - Fundação para a Ciência e a Tecnologia in the framework of the project PEst-C/SAU/LA0002/2013. T. Esteves is recipient of SFRH /BD / 80508/ 2011 by FCT. P. Quelhas is a *Ciência2008* awardee.

References

[1] A. Carvalho, A. Pelaez-Vargas, D. Gallego-Perez, L. Grenho, M.H. Fernandes, A.H. De Azaf, M.P. Ferraza, D.J. Hansford, and F.J. Monteiro. Micropatterned silica thin films with nanohydroxyapatite micro-aggregates for guided tissue regeneration. *Dental Materials*, 28(12):11, 2012.

[2] T. Esteves, M. J. Oliveira, and P. Quelhas. Cancer cell detection and tracking based on local interest point detectors. *Springer Lecture Notes in Computer Science, LNCS*, 7950:434–441, 2013.

[3] U. Vovk, F. Pernus, and B. Likar. A review of methods for correction of intensity inhomogeneity in mri. *IEEE Transactions on Medical Imaging*, 26(3):405–421, 2007.

Handwritten Signature Matching using GPUMLib

João Gonçalves[1]
jcgonc@dei.uc.pt
Noel Lopes[1,2]
noel@ipg.pt
Bernardete Ribeiro[1]
bribeiro@dei.uc.pt

[1] Department of Informatics Engineering
CISUC, University of Coimbra
Coimbra, Portugal

[2] Department of Informatics
Polytechnic of Guarda
Guarda, Portugal

Abstract

Today the programmable Graphics Processing Unit (GPU) has raised a noticeable interest for applications that demand high-computational power. In particular, biometric applications containing thousands of samples and features need efficient tools to process data. GPUMLib is an open source library with machine learning techniques endowed with GPU that is able to handle the significant memory and computational burden needed for signature matching. In this paper, the SVM component imbued with GPUMLib has been used for signature matching yielding good performance results assessed by the F-Score and False Positive Rate (FPR) in the GPDS database.

1 Signature Matching Background

Signature matching is a very important problem in authentication which covers a broad range of areas such as personal identification, security and bank transactions. Many efforts have been put to tackle the verification of signatures which contain biometric information. Often the databases are very large and such big data appears difficult to handle. Additionally, in offline settings, the lack of the dynamic characteristics makes the problem hard to solve.

This problem is very difficult for many reasons. The biometric data is a scanned 2D image. Unlike on-line verification there is lack of dynamic characteristics (e.g. velocity, pen pressure, acceleration, etc.) which reflect each individual motion style and are harder to fake. Additionally the biometric features in original and faked signatures can be extremely similar which makes the problem even harder. Examples are the shapes, sizes and variations of signatures that lead to a confluence of factors extremely tricky to verify. Also, the sheer volume of biometric data in many applications require fast tools for model selection in order to expose better models. Preprocessing of offline handwritten biometric data is complex and motivates the holistic study of many features capable of proper capturing the intra-variational characteristics of the individual signatures and the optimal group of features for building a better model.

Fast machine learning algorithms that are able to extract relevant information from large repositories play an important role. Therefore, we use in this work GPUMLib[1] which is an open-source machine learning library that will be described next.

2 GPUMLib

The GPUMLib framework is endowed with a wide range of machine learning algorithms implemented in CUDA using GPU. At its core, the library contains a set of CUDA kernels that support the execution of Machine Learning (ML) algorithms on the GPU. Usually, in order to implement an ML algorithm on the GPU several kernels are required. However, the same kernel might be used to implement different algorithms.

Each ML algorithm has its own C++ class that is responsible for: transferring the information (inputs) needed by the algorithm to the device (GPU); calling the algorithm kernels in the proper order; and transferring the algorithm outputs and intermediate values back to the host. This model allows non-GPU developers to take advantage of GPUMLib, without requiring them to understand the specific details of CUDA programming.

Moreover, GPUMLib provides a standard memory access framework to support the tasks of memory allocation and data transfer between the host and device (and vice-versa) in an effortless manner.

[1] http://gpumlib.sourceforge.net/

Figure 1: Genuine and forged signatures from the GPDS database.

3 Computational Experiments

In this section we describe the dataset, present the experimental setup and indicate the performance metrics for signature matching assessment.

3.1 Dataset

The database contains data from 300 individuals. For each individual there are 24 genuine signatures, plus 30 forgeries of his/her signature making 54 images per individual and a total of 16200 images. The 24 genuine specimens of each signer were collected in single day writing sessions. The forgeries were produced under the following conditions: The forger imitates a genuine signature from the static image of the genuine signature (scanned at 300 DPI) and the forger is allowed to practice writing the signature for as long as s/he wishes. Each forger has to imitate three signatures of five signers in a single day writing session. The genuine signature shown to each forger is chosen randomly from the 24 genuine ones. Therefore, for each genuine signature, there are 30 simple forgeries made by 10 forgers from 10 different genuine specimens. The dataset used consists of 16200 handwritten off-line signature recognition (each signature is a 649×462 pixels image). Additional information on this database can be found in Ferrer et al. [3].

3.2 Experimental Setup

For the test set we used 9 images and the remaining 45 for the training set. Both training and testing sets were randomly generated from the initial data, being the test set composed of 4 genuine signatures and 5 forged. The experiments were run 10 times per configuration.

With regard to feature extraction from the GPDS database previous research can be found in [1], [2] and [3]. From the aforementioned studies the best features so far were extracted from the original dataset. Following these authors and our previous work [5] in Table 1 we present the extracted features and the corresponding number of attributes from the GPDS image dataset.

3.3 Performance Metrics

We defined several measures based on the possible outcomes of the classification, namely, False Positive Rate ($FPR = \frac{FP}{FP+TN}$), and False Discovery Rate ($FDR = \frac{FP}{FP+TP}$), as well as combined measures, such as, the van Rijsbergen F_β measure, which combines recall ($R = \frac{TP}{TP+FN}$) and precision ($P = \frac{TP}{TP+FP}$) in a single score (F-Score $=$ F1 $= \frac{2PR}{P+R}$), yielding an harmonic average between precision and recall.

Table 1: Number of attributes of each feature.

Feature	Attributes
Best Fit	4
Discrete Cosine Transform (DCT)	5
Geometric Parameters (Cartesian)	180
Geometric Parameters (Polar)	192
Gravity Center	1
Histogram Frequencies (hist)	6
K-Means	10
Max Intensity Points (maxint)	1
Modified Direction Feature (MDF)	160
Six-fold-Surface	6
Three-fold-Surface	3
Wavelet Transform Feature	12

Table 2: NVIDIA GeForce 570 GTX characteristics.

Characteristic	Value
Number of SPs	480
IEEE single precision (float) performance	748.8 GFlops
Number of SMs	15
Shading clock speed	1.56 GHz
Memory size	1.25 GB
Memory bandwidth	152.0 GB/s
Shared memory per block	48 KB

4 Results and Discussion

In this section we describe the three experiments to deal with the signature matching problem using the SVM component of GPUMLib [4].

The Experience 1 comprises the identification of original and forged signatures. For that purpose, we used all the 300 individuals and studied each group of features. The best feature combination DCT + MDF was tested for all the supported kernel functions by our GPU-based SVM Classifier [5]. The results obtained by 5-Fold cross-validation are shown in Figure 2.

In the second experience[2], instead of using all the 300 groups of signatures we exploited several groups or combinations of features for each individual. Therefore, this experiment consisted of identifying, for each person, if a signature was either original or forged. Only the Radial Basis Function (RBF) kernel was employed. For validation the test set consisted of genuine 4 original plus 5 forged signature, as specified above.

In the Experience 3, we performed signature matching related to a given individual, using a One-Against-One multi-class classifier, that is, we trained and tested each individual class against one of the others. As the cost involved in the training process is high, we only used the RBF kernel and 5 K-Fold cross-validation procedure. The results are illustrated in Figure 3.

5 Conclusions

Handwritten signature matching plays a crucial role in many important transactions for security and privacy reasons. In this work, we presented the performance analysis of the SVM component of the GPUMLib which yielded good results in the GPDS database. To this end, several experiments were performed using the features extracted in a previous work. One possible direction of future work is to integrate multiple kernel learning in the GPUMLib for this kind of pattern recognition problems.

References

[1] S. Armand, M. Blumenstein, and V. Muthukkumarasamy. Off-line signature verification using an enhanced modified direction feature with single and multi-classifier approaches. *Computational Intelligence Magazine, IEEE*, 2(2):18 –25, 2007.

[2]Results from Experience 2 are omitted due to paper's space restrictions.

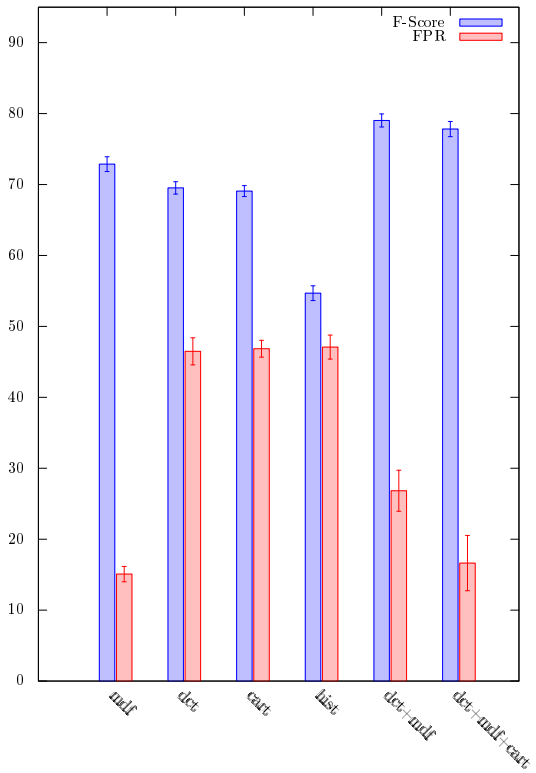


Figure 2: Experience 1: Results with the RBF kernel.

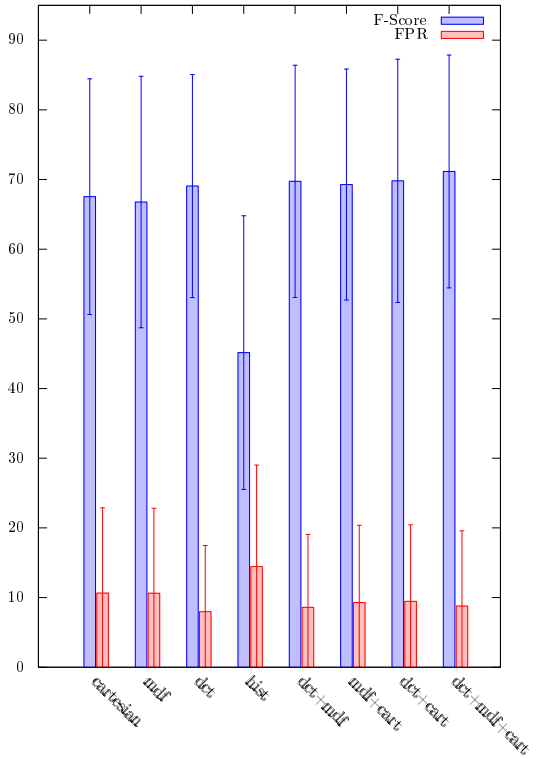


Figure 3: Experience 3: Forged/original signature per individual.

[2] M. Blumenstein, X.Y. Liu, and B. Verma. A modified direction feature for cursive character recognition. In *Proc of IEEE Int J Conf on Neural Networks*, volume 4, pages 2983–2987, 2004.

[3] M. Ferrer, J. Alonso, and C. Travieso. Offline geometric parameters for automatic signature verification using fixed-point arithmetic. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 27 (6):993–997, 2005.

[4] Noel Lopes and Bernardete Ribeiro. GPUMLib: An efficient open-source GPU machine learning library. *Int J of Computer Information Systems and Industrial Management Applications*, 3:355–362, 2011.

[5] Bernardete Ribeiro, Noel Lopes, and João Gonçalves. Signature identification via efficient feature selection and GPU-based SVM classifier. In *Proc. of IEEE Int J Conf on Neural Networks (IJCNN)*, 2014.

Border Abruptness Assessment Methodology for Skin Lesion Image Analysis

Maria João M. Vasconcelos, Luís Rosado
{maria.vasconcelos, luis.rosado}@fraunhofer.pt
Fraunhofer Portugal AICOS,
Porto, Portugal

Márcia Ferreira
Portuguese Institute of Oncology,
Porto, Portugal

Abstract

A methodology to assess the border abruptness of a skin lesion is presented, starting to extract 97 abruptness features, followed by testing 4 feature selection methods and 5 machine learning classification algorithms to maximize the methodology performance. Two datasets with different image types, dermoscopic and mobile acquired images, previously annotated by specialists, were used. For dermoscopic images, the developed methodology achieves an accuracy of 81% regarding border segment classification, while for the mobile acquired images the accuracy reaches 74.8%.

1 Introduction

In Portugal, every year around 700 new cases of melanoma and 10.000 new cases of nonmelanoma are diagnosed [1]. In addition, skin cancer represents the most frequent cancer type nowadays. Even though Malignant Melanoma (MM) accounts for only a small percentage of skin cancer, it is far more dangerous than others and it is responsible for most skin cancer related deaths.

Prevention is essential to fight back the previous stated facts, so, recently many efforts have been made to design accurate computer aided diagnosis (CAD) systems capable of automatically analyse skin lesions, being most of them designed for dermoscopic images. In addition, the spreading of the new generation of mobile devices, with built-in image cameras with remarkable resolution, opens a window for the development of mobile-based approaches for initial diagnosis and screening of skin lesions.

The most common criteria that clinicians and dermatologists follow to determine the risk of a given lesion being a melanoma is the ABCD rule of dermatoscopy [2]. It analyses four visual features of the skin lesion: asymmetry, border, colour and differential structures; then, the Total Dermatoscopy Score (TDS) is given by the multiplication of each feature score by a weight factor (1.3, 0.1, 0.5 and 0.5, respectively) and, depending on the value obtained, a grading of the lesion is possible with respect to their malignant potential. According to the ABCD rule, the border is divided into 8 segments (see Figure 1a) and the score ranges from 0 to 8, scoring 1 for the presence of each abrupt border segment.

In this work, a methodology to assess the border abruptness score of a skin lesion image, as proposed by the ABCD rule, is presented.

2 Datasets

The methodology presented in this work was applied to two distinct image types: dermoscopic and mobile acquired images.

The dermoscopic image set consists on images obtained from the EDRA Interactive Atlas of Dermoscopy [3]. From all images of the dataset, only the ones with a complete skin lesion visible were considered, giving a total of 298 digital images.

The other dataset used was collected at the Portuguese Institute of Oncology of Porto, under the scope of the project Melanoma Detection [4]. The dataset contains 80 24-bit colour images with 652x652 pixels of resolution, acquired with a HTC One S mobile phone.

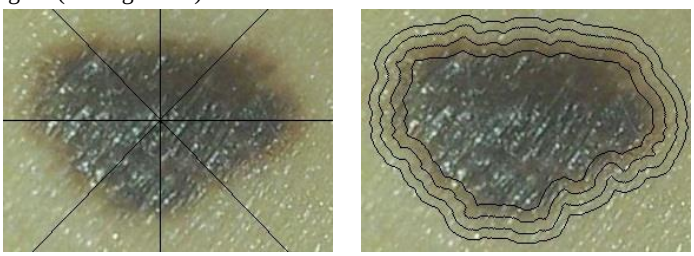
All images from both datasets were annotated by medical experts regarding the border score of the lesion, as proposed by the ABCD rule. Table 1 summarizes the score classification for both datasets.

3 Methodology

The first step of this methodology consists in performing the lesion segmentation. For that, an adaptive segmentation algorithm was used, based on the Otsu's threshold method [5]. The adaptive algorithm considers the threshold value as the mean intensity value of the squared region centred each the image pixel.

Then, the axes of the lesion are calculated as described in [6], by considering the first axis of inertia as the primary axis, while the secondary axis is perpendicular to the primary axis and passes through the centroid of the lesion. The lesion is later divided in eight equal parts as depicted in Figure 1a.

Afterwards, and in order to overcome possible border segmentation errors, three distinct regions are considered: inner border, border and outer border regions. As suggested in [6], a region inside the border with an area of 10% of the lesion and outside the border with the same area is considered as the border region, in addition two other regions with 20% of the area of the lesion were extracted, inside and outside the border region (see Figure 1b).



(a) (b)

Figure 1 - Example of the border segments division (a) and inner, border and outer regions considered (b).

Table 1 – Doctor score of the border abruptness of the lesions.

Score	0	1	2	3	4	5	6	7	8	#
EDRA	83	8	32	40	37	20	27	0	51	298
Mobile	25	37	10	5	2	0	1	0	0	80

A total of 97 features were calculated for each segment of the lesion as summarized in Table 2 and briefly explained next.

Features of colour and binary groups were proposed by other researchers, while the here called radial features combined existing principles in order to create a new approach. The images were also transformed from the original RGB colour space to the HSV and CIE Lab colour space, being the R, G, B, V and L channels used to extract border related features.

For the colour group, statistics like mean and standard deviation of the inner, border and outer region were calculated for each indicated channel and considered as individual features [7]. The meanratio feature is obtained by subtracting the maximum by the minimum mean values of the inner and outer region and dividing it by the maximum value; the stdratio corresponds to the division of the standard deviation by the mean value of the border region. Moreover, for each channel, the magnitude of filtered image using the Sobel operator for the horizontal and vertical directions was calculated, and the corresponding mean and standard deviation obtained (meangrad and stdgrad) [7,8].

For the radial group, instead of considering all pixels from each area to obtain the features, the radial features use only the pixels belonging to the lines that divide each segment region in 9 equal parts, or by other words, to the lines that pass through the lesion centroid and make an angle multiple of 5 degrees with the principal axis. With this procedure, the ones intend to compare corresponding values in each inner, border and outer region to deduce the abruptness level of the border. Statistics such as mean, standard deviation, mean ratio, standard deviation ratio and mean slope [9] of the pixels intensities were obtained and considered as individual features.

Finally, in the binary group, features like solidity and compactness index were also obtained from the binary mask of each segmented region [6,8].

For classification purposes, since the border annotations refer to the whole lesion, and no indication exists of which border segment is considered abrupt (1) or smooth (0), the features were ordered according to their meanratio value in the blue channel, and score 1 was attributed to border segments with higher values.

Table 2 – Summary of the extracted border features.

Group	Channels	Features	Refs.
Colour	R, G, B, H, V	mean and std (regions: inner, border, outer), meanratio, stdratio, meangrad (border), stdgrad (border)	[7,8]
Radial	R, G, B, H, V	mean and std (inner, border, outer), meanratio, stdratio, meanslope	[7,8, 9]
Binary	Binary	solidity, compactness index	[6,8]

Feature selection was applied in order to improve classification results and four different methods were chosen: two filter-based and two wrappers. The filter-based methods used were the Correlation based Feature Selection (CFS) and the ReliefF; and search methods used for wrapper were Linear Forward Selection and Greedy.

Five different classifiers were adopted in the training process, namely: Naïve Bayes, Support Vector Machines (Sequential Minimal Optimization), K-Nearest Neighbours (with k equal to 3), Boosting (Adaboost M1) and Random Tree. Moreover, in the feature selection step, the wrapper methods were performed independently for each classifier.

4 Results

The suggest method was developed using C++ and OpenCV library [10] and classification results were obtained with Weka software [11].

After extracting all features, performing feature extraction and classification with the various methods, statistics like sensitivity, specificity and accuracy were obtained to infer about the best combination for each dataset. For the EDRA dataset, 10-fold cross validation was used for the 2384 segments classification, corresponding to the 298 dermoscopic images. The low number of images in the Mobile dataset coupled with the low percentage of borders considered abrupt by the doctor, resulted in a very unbalanced set to classify. So, a training set with 60% of all border segments that were considered abrupt was selected and the same number of smooth segments, resulting on a training set of 52 segments of each class (1- abrupt and 0-smooth). The remaining 536 segments were used to test the classifiers.

For the EDRA dataset, the best classification results were obtained by the 3-nearest neighbour classifier when using the wrapper method with LFS as search method. A total of 13 features were chosen by the selection procedure, where 7 belong to the colour group and 3 of them were extracted from the gradient images and the other 6 features belong to the radial group, with 3 being from meanslope statistic.

The selected combination of features for the dermoscopic dataset achieved a sensitivity rate of 75.5%, a specificity of 84.9% and an accuracy of 81%, with the confusion matrix being depicted in Table 3. After grouping the results for each segmented region into the corresponding image lesion score (see Table 4), it is possible to observe that the automatic classification shows promising results when compared with the doctor classification, considering that is a 9-class classification problem.

For the Mobile dataset, Naïve Bayes classifier using the wrapper approach with LFS as search method achieved the best performance. The selected features were meanratio, meangrad, radialmeanstdratio and meanIn from R, L, B and R channels, respectively.

The confusion matrix for the selected features is shown in Table 3, what translates in 67.6% of sensitivity, 75.3% of specificity and an accuracy of 74.8%. Given the low percentage of borders considered abrupt by the doctor in the Mobile dataset, it was not possible to obtain the classification results for each lesion. Despite this, and considering the difficulty of analysing this type of images, 74.8% of accuracy can be considered quite satisfactory for our purposes.

Table 3 – Confusion matrix of the best classifier for each dataset.

		Doctor classification	
Aut om.		1	0
	1	763	207
	0	247	1167

EDRA

		Doctor classification	
Aut om.		1	0
	1	23	124
	0	11	378

Mobile

Table 4 – Confusion matrix of the best classifier considering each lesion of the EDRA dataset.

		Doctor classification								
		0	1	2	3	4	5	6	7	8
Automatic Classification	0	53	3	4	2	-	-	-	-	1
	1	18	2	9	9	14	1	1	-	-
	2	7	3	8	6	2	2	-	-	1
	3	4	-	4	13	8	-	1	-	3
	4	1	-	6	5	8	4	1	-	3
	5	-	-	1	2	6	7	5	-	4
	6	-	-	-	2	6	4	7	-	6
	7	-	-	-	1	1	1	11	-	12
	8	-	-	-	-	2	1	1	-	21

5 Conclusion

In this work, a methodology to assess the border score of a skin lesion based on the analysis of the image is proposed. While most of the approaches proposed to date classify the lesion as a whole, this work focus in the binary classification of each border segment and only after infers about total score of the lesion, regarding the border abruptness classification.

Two images types, dermoscopic and mobile acquired, are considered with 81% and 74.8% of accuracy rates achieved in the classification of border segments, respectively.

Acknowledgements

This work was done under the scope of the project “SMARTSKINS: A Novel Framework for Supervised Mobile Assessment and Risk Triage of Skin Lesion via Non-invasive Screening” with reference PTDC/BBB-BMD/3088/2012 financially supported by Fundação para a Ciência e a Tecnologia in Portugal.

References

[1] Liga Portuguesa Contra o Cancro [Online]. http://www.ligacontracancro.pt/.

[2] W. Stolz, A. Riemann, et. al. ABCD rule of dermatoscopy – a new practical method for early recognition of malignant melanoma. European Journal of Dermatology 4: 521-527, 1994.

[2] G. Argenziano, H.P. Soyer, et. al. Dermoscopy: a tutorial. Milan. EDRA Medical Publishing & New Media. 2002.

[4] P.A. Fraunhofer. Melanoma detection, internal project. http://www.fraunhofer.pt/en/fraunhofer_aicos/projects/internal_research/melanoma_detection.html, 2014.

[5] N. Otsu. A threshold selection method from gray-level histograms. IEEE Transactions on Systems, Man and Cybernetics, 9(1): 62-66, 1979

[6] M.E. Celebi, H.A. Kingravi, et al. A methodological approach to the classification of dermoscopy images. *Computerized Medical Imaging and Graphics* 31: 362-373, 2007.

[7] P.G. Cavalcanti, J. Scharcanski. Automated prescreening of pigmented skin lesions using standard cameras. *Computerized Medical Imaging and Graphics* 35(6): 481-491, 2011.

[8] J.F. Alcon C. Ciuhu, et. al. Automatic imaging system with decision support for inspection of pigmented skin lesions and melanoma diagnosis. IEEE Journal Select Top Signal Process 3: 14-25, 2009.

[9] G.R. Day. How blurry is that border? An investigation into algorithmic reproduction of skin lesion border cut-off. *Computerized Medical Imaging and Graphics* 24: 69-72, 2000.

[10] G. Bradski and A. Kaehler. Learning OpenCV: Computer Vision with the OpenCV Library. O’Reilly Media Inc, 2008.

[11] M.Hall, E. Frank, et. al. The WEKA data mining software: An update. SIGKDD Explorations 11(1): 10-18, 2009.

Classification of Lung Function on a Smartphone App

João Teixeira[1]
jpftexas@gmail.com
Luís Teixeira[2]
luisft@fe.up.pt
João Fonseca[3]
jfonseca@med.up.pt
Tiago Jacinto[3]
tajacinto@gmail.com

[1] Center for Health Technology and Services Research, Department of Electrical and Computer Engineering University of Porto, PT
[2] Department of Informatics Engineering University of Porto, PT
[3] Center for Health Technology and Services Research, Department of Health Information and Decision Sciences University of Porto, PT

Abstract

Asthma and COPD are chronic lung conditions that affect over 250 million people worldwide. They can result in breathlessness, a harsh decrease in quality of life and even death, if not detected and duly managed. In this paper, we aim to find the best and most efficient combination of signal processing and machine learning methods to produce a smartphone app that could accurately classify lung function, using microphone recordings as the only input. A dataset of 101 recordings of forced expiration maneuvers from 61 patients was gathered. The signal processing comparison experiments were conducted in a backward selection approach, reducing from 54 to 12 final envelopes, per recording. The classification experiments focused first on differentiating Normal from Abnormal lung function, and second in multiple lung function patterns.

1 Introduction

Chronic respiratory diseases such as Asthma and Chronic Obstructive Pulmonary Disease (COPD) are incurable, yet treatable and their early detection is crucial to provide a better quality of life. Major risk factors include air pollution, tobacco smoking and occupational environments containing dust and chemicals. The World Health Organization (WHO) estimates that over 250 million people suffer from asthma and COPD and 3 million people died of COPD in 2005.

The increasing use of smartphones has enabled the emergence of several health related systems. Their computational power is ever increasing and, equipped with multiple sensors, it is possible to develop disease prevention, diagnosis and monitoring applications.

The aim of this paper is to compare several groups of methods and clinical parameters in order to find the most relevant, most efficient and faster combination to produce a smartphone app for measuring and classifying lung function. The system's input is restricted to the smartphone's built-in microphone, in order to avoid external components.

2 Background and Related Work

Traditional spirometers accurately measure a wide range of lung function parameters but have the disadvantage of being very expensive and being of sizable dimensions. Portable spirometers were developed to meet the needs of home spirometry and thus, they disregard many unused functions and measurements. Additionally, their cost and dimensions are smaller. The most recent spirometers use laptops as computational platforms and airflow sensors that use USB connections. This enabled to reuse the computing platform and easily transport the system. The next logical step concerning portability and affordability involves lung function estimation with smartphones' microphones. Some studies have already been conducted in order to accurately measure the clinical parameters [1, 4, 5].

3 Data Collection Procedure

The dataset is composed by 101 recordings from 61 caucasian adult patients performing the forced expiration maneuver, without any mouthpiece, at an arm's length. Each recording was accompanied by the patient's anthropometric data (age, height, weight and gender), clinical parameters, and classification of the patient's lung function provided by the recording physician. The classification types are normal and abnormal (obstruction, restriction or mixed). The included patients were either part of the Control and Burden of Asthma and Rhinitis (ICAR) clinical study, where lung function was measured, or were attending allergology consultations at CUF Porto Institute or CUF Porto Hospital.

4 Algorithms and System Architecture

4.1 Signal Processing

The audio input was initially segmented in order to remove non expiration sounds, using both back-extrapolation [3] and sliding window maximum ratio detection algorithms. Then, the signal was converted to pressure at the lips, p_{lips}, through an Inverse Radiation Model, which was afterwards converted to airflow at the lips, u_{lips}, using a Pressure to Flow Conversion Model, similarly to Larson *et al*. [1].

This second stage employed several methods to calculate the signal envelopes. Such methods include the Hilbert Transform, Linear Predictive Coding (LPC), orders 2, 4, 8, 16, and 32, Mean of Resonances, all of which described on Larson *et al*. [1], and Shannon Entropy and Energy envelopes described on Liang *et al*. [2].

On the final stage, the envelopes are filtered or smoothed with a Low Pass Filter (LPF), Moving Average and a Savitzky-Golay polynomial filter (SG). The unprocessed envelopes are also approximated by a polynomial function.

4.2 Parameter Extraction

For each recording, the spirometry parameters were calculated from each of the final envelopes. The measurements extracted were PEF, FVC, FEV_1, FEV_1/FVC, $FEF_{25\%-75\%}$, $FEF_{25\%}$, $FEF_{50\%}$, $FEF_{75\%}$ and a custom parameter proposed in Stein [4]. The envelopes are viewed as Flow-Time curves, characteristic of spirometer reports.

PEF is defined as the Peak Expiratory Flow or the global maximum of the audio envelope. By integrating the envelope with respect to time the Volume-Time curve can be obtained. FVC is defined as the total volume expired of a FEM. FEV_1 is the total volume expired during the first second. $FEF_{25\%-75\%}$ corresponds to $^1/_2$FVC / $(t_{75\%} - t_{25\%})$, in which $t_{x\%}$ is the time at which the volume corresponds to $x\%$ of the FVC. $FEF_{x\%}$ is the instantaneous flow value at x% of the total volume.

4.3 Machine Learning

For the clinical parameters' regression Tree Bagging and Random Forests were used, while spirometer measurements acted as ground truth or regression target. Using the regressed clinical parameters, several classification models were devised. The models used were Decision Trees, either as one tree, Tree Bagging, Random Forest and AdaBoost, Support Vector Machines (SVM) and Naïve Bayes.

5 Experimental Approach and Results

5.1 Parameter Regression Experiments

The algorithms used on the experiments were based on a backward selection approach. Initially, all the signal processing methods were used and the clinical parameters' sets were obtained by successively removing some methods out of the initial set. For each set, a 5-fold cross validation

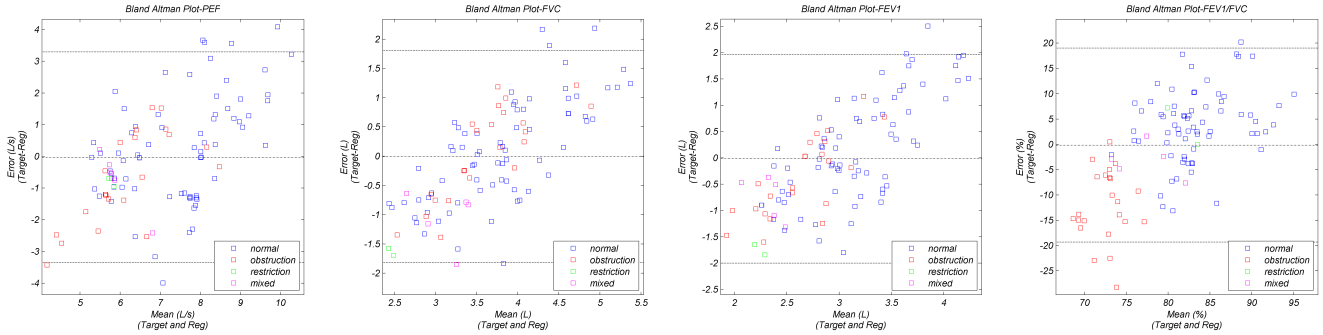


Figure 1: Bland Altman plots of the errors between the spirometer target and the regressed values versus the mean between the target and regressed values. The mean (dashdot) and $\pm 2\sigma$ (dash) indicative lines are also shown.

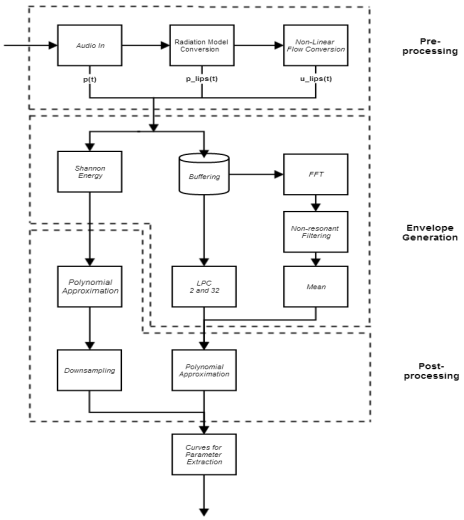


Figure 2: Final Signal Processing Architecture

set was made to verify the models' expected accuracy and to obtain the average regression error and standard deviation. The method comparison experiments included pre and post-processing filtering options and redundancy verification of some envelope generating functions. Two types of measurements were used to assess the method sets, the simple comparison of the regression average error and parameter tree split count and distribution which represents the gross importance of each envelope process. An envelope process is, for instance, $p \rightarrow p_{lips} \rightarrow$ LPC 32 $\rightarrow$ Polynomial Approximation, as represented on Figure 2.

Almost all removed methods did not induce a significant change on the regression results. When given two alternative method that did not change the results the choice of which to discard was based on the relative importance presented by the split count test, or based on computational complexity arguments. Figure 2 presents the final signal processing architecture based on the regression experiments of the methods studied. Throughout the experiments, regressing the parameters using Bagging and Random Forest always presented very similar results. Therefore, Random Forest was chosen for regressing the parameters for the classification experiments, reducing overfitting and regression time. The Bland-Altman analysis of the regressed parameters revealed that are the models are valid, not very accurate on most parameters, and tend to over-estimate the spirometers measurements of the lower volume or flow samples.

5.2 Classification Experiments

The classification process was first devised as a Normal against Abnormal classification problem, referred to as two label experiments (TLE). Then, multiple label experiments (MLE) were conducted, where the models tried to distinguish between Normal lung function and Obstruction, Restriction and Mixed pathologies. For both problems, the experiments varied on the feature space used. They compared different combinations of sets A[1] and B[2], and the anthropometric values height and age.

Generally of TLE and MLE, the tree ensemble methods and Linear SVM presented the best results. For simplicity of implementation and speed of training and testing, Random Forest was chosen for the final prototype. For both problems, incrementing the patients' height and age and removing Set B improved the results across the learning methods. This results are consistent with the way physicians evaluate lung function.

6 Conclusion

The regression experiments enabled to reduce the 54 to 12 final envelopes, per recording, without a significant change on the results. The regression mean error for the less error prone parameters was 21%, 30%, 24% and 9.7%, for PEF, FEV_1, FVC, and FEV_1/FVC, respectively. The classification models obtained an error rate of 8% and 17%, for the TLE and MLE, respectively. Also, the TLE model presented a precision of 90.91% and recall of 98.59%. In conclusion, our work demonstrated that smartphone spirometry and automatic lung function triage is possible and the results encourage further development of the systems.

References

[1] Eric C. Larson, Mayank Goel, Gaetano Boriello, Sonya Heltshe, Margaret Rosenfeld, and Shwetak N. Patel. SpiroSmart: Using a Microphone to Measure Lung Function on a Mobile Phone. In *14th ACM International Conference on Ubiquitous Computing*, page 10, Pittsburgh, Pennsylvania, USA, 2012. ISBN 9781450312240.

[2] H Liang, S Lukkarinen, and I Hartimo. Heart sound segmentation algorithm based on heart sound envelogram. In *Computers in Cardiology 1997*, volume 24, pages 105–108. IEEE, 1997. ISBN 0-7803-4445-6. doi: 10.1109/CIC.1997.647841. URL `http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=647841`.

[3] M R Miller, J Hankinson, V Brusasco, F Burgos, R Casaburi, A Coates, R Crapo, P Enright, C P M van der Grinten, P Gustafsson, R Jensen, D C Johnson, N MacIntyre, R McKay, D Navajas, O F Pedersen, R Pellegrino, G Viegi, and J Wanger. Standardisation of spirometry. *The European respiratory journal*, 26(2):319–38, August 2005. ISSN 0903-1936. doi: 10.1183/09031936.05.00034805. URL `http://www.ncbi.nlm.nih.gov/pubmed/16055882`.

[4] Bas van Stein. A Mobile Smart Care platform Home spirometry by using the smartphone microphone. Master's thesis, Leiden University, Leiden, The Netherlands, 2013.

[5] Wenyao Xu, Ming-chun Huang, Jason J. Liu, Fengbo Ren, Xinchen Shen, Xiao Liu, and Majid Sarrafzadeh. mCOPD. In *Proceedings of the 6th International Conference on Pervasive Technologies Related to Assistive Environments - PETRA '13*, PETRA '13, pages 1–8, New York, New York, USA, 2013. ACM Press. ISBN 9781450319737. doi: 10.1145/2504335.2504383. URL `http://doi.acm.org/10.1145/2504335.2504383http://dl.acm.org/citation.cfm?doid=2504335.2504383`.

[1] Set A: PEF, FEV_1, FVC, FEV_1/FVC

[2] Set B: $FEF_{25\%-75\%}$, $FEF_{25\%-75\%}/FEV_1$, $FEF_{25\%}$, $FEF_{50\%}$ and $FEF_{75\%}$

Blind Inpainting after Log Transformation using Total Variation

Manya V. Afonso
http://mafonso.weebly.com
João M. R. Sanches
http://users.ist.isr.utl.pt/~jmrs

Instituto de Sistemas e Robótica
Instituto Superior Técnico
Lisboa, Portugal.

Abstract

In this paper, we propose a method for blind image inpainting, in which the locations of the pixels which are missing or corrupted with impulse noise are not known. A logarithmic transformation is applied to convert the multiplication between the image and binary mask into an additive problem. The image and mask terms are then estimated iteratively with total variation regularization applied on the image, and ℓ_0 regularization on the mask term. The resulting alternating minimization scheme simultaneously estimates the image and mask, in the same iterative process. Experimental results show that the proposed method can deal with a larger fraction of missing pixels than two phase methods which first estimate the mask and then reconstruct the image.

1 Introduction

It is known that faulty sensors or bit errors during transmission can cause some pixels in an image to be lost or corrupted by impulse noise [5, 10]. In the case of missing pixel values, the corrupted pixels are assumed to have a value equal to zero, and the problem of estimating the complete image is called the inpainting problem.

We represent the image to be estimated with n pixels as a vector, say in lexicographic ordering, $\mathbf{x} \in \mathbb{R}^n$. Let $m < n$ be the number of observed pixels or pixels free from impulse noise. The mapping from $\mathbf{x}$ to the partially observed image $\mathbf{y}$ is given by,

$$\mathbf{y} = \mathbf{A}\left(\mathbf{x} + \mathbf{n}_g\right), \tag{1}$$

where the mask operator $\mathbf{A}$ is a $n \times n$ identity matrix with the diagonal elements equal to zero, corresponding to missing pixels, and $\mathbf{n}_g$ is an additive Gaussian noise term corrupting the observed pixel values.

When the index set of the observed pixels or the observation mask $\mathbf{A}$ is known, the inpainting problem can be solved by one of several existing methods for image reconstruction from a sparse set of observations [1, 2].

Adaptive filtering [10], [9] can be used to detect outliers and remove impulse noise for low levels of missing/damaged pixels. Some two phase methods such as [6] have a mask detection step in which the mask is detected through the use of one of these filters and the image is estimated in a separate step.

1.1 Contributions

In this paper, we propose a method to estimate the image $\mathbf{x}$ without knowing the observation mask $\mathbf{A}$ apriori, *i.e.*, we simultaneously estimate the image and the mask. We formulate the masking operation as a summation after logarithmic compression, and apply a total variation (TV) [8, 12] regularizer on the image term, and an ℓ_0-norm regularizer on the term corresponding to the mask. The TV regularizer encourages the estimate of $\mathbf{x}$ to be piece-wise smooth, while the ℓ_0-norm regularizer encourages the mask term to be sparse. Experimental results show that our proposed method can deal with as many as 95% of the pixels missing, which is higher than reported in literature.

2 Proposed Method

Given the observation $\mathbf{y}$, our goal is to estimate the image $\mathbf{x}$ and the mask $\mathbf{A}$. Since $\mathbf{A}$ is a diagonal matrix $\mathbf{A} = \text{diag}(\mathbf{a})$ and the masking operation is element-wise multiplication, we will optimize over the vector of diagonal elements $\mathbf{a} \in \{0,1\}^n$. When a pixel with index i is observed, the corresponding mask element $a_i = 1$, and when pixel i is lost, $a_i = 0$. Thus, a pixel k in vector $\mathbf{y}$ is defined as, $y_i = x_i \times a_i$.

We do not know apriori if a given pixel y_k corresponds to an observed one ($a_k = 1$) or not ($a_k = 0$). Rather than have $a_i = 0$ when the pixel is not observed, we can define a_i to be a small value in the order of 10^{-K} or smaller, K being a positive integer greater than or equal to 3. Defining $v_i = \log(a_i)$, we have,

$$v_i = \begin{cases} 0, & \text{if } i \text{ is observed} \\ -K, & \text{otherwise} \end{cases} \tag{2}$$

If the maximum possible pixel value is 255, the value of K must satisfy $K > \log 255$. Assuming that $\mathbf{y}$ and $\mathbf{x}$ are always positive, applying a logarithmic transformation on (1) converts it into an additive model, $g_i = u_i + v_i$, where $u_i = \log x_i$, and $g_i = \log(y_i + \delta)$. A small positive bias term $\delta > 0$ is added to $\mathbf{y}$ to guarantee positivity.

Our problem is now estimating the vectors $\mathbf{u}$ and $\mathbf{v}$, given the log transformed observation $\mathbf{g}$. We assume that our image $\mathbf{x}$, and therefore its logarithmic transformation $\mathbf{u}$ are piece-wise smooth. We therefore apply a TV regularizer on the log transformed image $\mathbf{u}$, and the ℓ_0-norm regularizer on the log-transformed mask $\mathbf{v}$. Our estimation problem is therefore,

$$(\hat{\mathbf{u}}, \hat{\mathbf{v}}) = \arg\min_{\mathbf{u},\mathbf{v}} \frac{1}{2}\|\mathbf{g} - \mathbf{u} - \mathbf{v}\|_2^2 + \frac{\lambda_1}{2} TV(\mathbf{u}) + \frac{\lambda_2}{2}\|\mathbf{v}\|_0, \tag{3}$$

where $\lambda_1, \lambda_2 > 0$ are the respective regularization parameters.

Formulations of the form (3) with different regularizers have appeared in the context of image decomposition [13] and deblurring with a sum of regularizers [3]. Alternating minimization schemes involving a sum of the ℓ_0-norm and a convex term were also used in [4, 11] for sparse image recovery. Since (3) is a separable problem, we can apply a simple iterative alternating method as in [3].

Isolating the terms in each variable keeping the other fixed, leads to a Gauss-Seidel scheme. Solving for $\mathbf{u}$ at iteration t,

$$\hat{\mathbf{u}}^{(t)} = \arg\min_{\mathbf{u}} \frac{1}{2}\|\mathbf{g} - \mathbf{u} - \mathbf{v}^{(t)}\|_2^2 + \frac{\lambda_1}{2} TV(\mathbf{u}). \tag{4}$$

This is a TV regularized denoising problem, the solution of which can be computed efficiently using Chambolle's algorithm [7].

Similarly, for $\mathbf{v}$ at iteration t we have,

$$\hat{\mathbf{v}}^{(t)} = \arg\min_{\mathbf{u}} \frac{1}{2}\|\mathbf{g} - \mathbf{u}^{(t)} - \mathbf{v}\|_2^2 + \frac{\lambda_2}{2}\|\mathbf{v}\|_0. \tag{5}$$

This problem although non-convex, has a solution given by the hard threshold,

$$\hat{\mathbf{v}}^{(t)} = H_{\lambda_2}(\mathbf{g} - \mathbf{u}^{(t)}), \tag{6}$$

where $H_{\lambda_2}(.)$ is the hard threshold operator and is defined element-wise as,

$$v_i^{(t)} = \begin{cases} 0, & \text{if } (g_i - u_i^{(t)}) \leq \lambda_2, \\ (g_i - u_i^{(t)}), & \text{otherwise.} \end{cases} \tag{7}$$

This iterative process is run until the relative difference between successive estimates falls below a given threshold. The estimates of the image and mask are computed by inverting the logarithmic transformation, $\hat{\mathbf{x}} = 10^{\hat{\mathbf{u}}}$ and $\hat{\mathbf{a}} = 10^{\hat{\mathbf{v}}}$.

3 Experimental Results

All experiments were performed on MATLAB on an Ubuntu Linux based server with 64 GB of RAM.To test our proposed method, we generate a random binary mask with a fraction of its elements equal to zero and multiply it element-wise to our image corrupted with additive Gaussian noise. The criteria used to evaluate the accuracy of estimation are the Improvement in Signal to Noise Ratio (ISNR), which is defined as $10\log_{10}\left(\frac{\|\mathbf{y}-\mathbf{x}\|^2}{\|\mathbf{y}-\hat{\mathbf{x}}\|^2}\right)$,

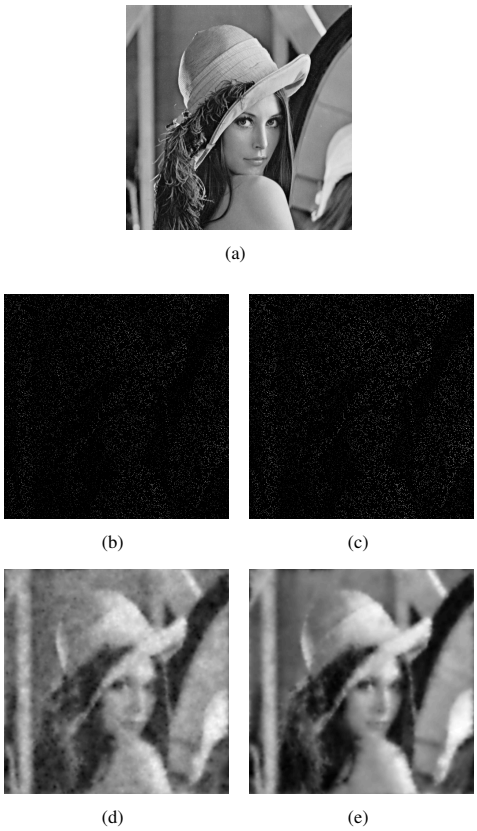


Figure 1: Inpainting with the Lena image: (a) original image, (b) observed with Gaussian noise (SNR 5 dB) and 95% pixels missing, (c) observed with Gaussian noise (SNR 20 dB) and 95% pixels missing, (d) estimate of (b) using the proposed method. ISNR = 15.12 dB, SSIM = 0.634, (e) estimate of (c) using the proposed method. ISNR = 15.89 dB, SSIM = 0.689.

the structural similarity index measure (SSIM) [15], and the fraction of incorrectly estimated mask pixels. The latter is computed using the binary exclusive or (XOR) operation with the original mask.

The iterative process was run until the relative difference between successive iterates $\|\hat{\mathbf{x}}^{(t+1)} - \hat{\mathbf{x}}^{(t)}\| / \|\hat{\mathbf{x}}^{(t+1)}\|$ fell below a threshold of 10^{-3}. In the results presented in this paper, the values of the regularization parameters used were $\lambda_1 = 0.008$ and $\lambda_2 = 0.1$, which were found to work well. Figure 1 shows the worst case results obtained with the proposed method for the 512×512 Lena image. The binary mask has only 5% of its pixels equal to 1, which means that 95% of the pixels were randomly discarded. Figures 1(b) and 1(c) show the observed images obtained with this mask, with additive Gaussian noise with Signal to Noise Ratios (SNR) of 5 dB and 20 dB, respectively.

Table 1 compares the proposed method with other methods for blind inpainting, for Gaussian noise with $\sigma = 15$. The fraction of missing pixels is varied from 0.1 to 0.9. The methods, fast two phase deblurring with TV [6], Adaptive Outlier Pursuit (AOP) [16], and K-ALS [14] were run for the ranges of fractions of missing pixels, as reported by their authors. We can see from these tables that [16] and [14] produce better ISNRs with fewer missing pixels, but our proposed method produces a better ISNR for higher percentages of missing pixels. While [6] is more accurate in estimating the observation mask, it produces a lower ISNR.

4 Conclusions

We have presented an iterative method for inpainting an image without knowing the locations of the missing pixels, based on alternating minimization to simultaneously estimate the image and observation mask. The method can deal better with a larger fraction of missing pixels than existing methods. Preliminary results show that the method is promising for the removal of impulse noise. Current and future work include optimizing the regularization parameters, a detailed convergence analysis, and extension of the framework to non-Gaussian observation models.

Table 1: Comparison for blind inpainting, with the Lena image, $\sigma = 15$ (SNR = 10.1 dB).

Fraction missing	Method	ISNR (dB)	CPU time (sec)	SSIM	Mask errors (%)
0.1	Proposed	15.03	**62.31**	**0.912**	**0.032**
	AOP	12.04	94.76	0.795	1.45
	KALS	**16.13**	378.19	0.893	1.54
0.25	Proposed	**18.72**	**53.1**	**0.906**	0.053
	[6]	1.29	264.09	0.392	**0.016**
	AOP	15.23	145.45	0.778	3.93
	KALS	16.71	377.94	0.846	4.62
0.4	Proposed	**19.90**	**53.46**	**0.895**	0.042
	[6]	5.49	265.88	0.43	**0.018**
	AOP	15.69	204.07	0.759	7.12
	KALS	7.06	370.87	0.464	19.40
0.5	Proposed	**19.98**	**58.78**	**0.882**	**0.02**
	[6]	4.98	900.09	0.385	4.15
	AOP	10.19	319.36	0.643	12.37
0.7	Proposed	19.49	63.14	0.842	0.0088
0.9	Proposed	17.22	92.29	0.738	0.0038

Acknowledgements

This work was supported by *Fundação para a Ciência e Tecnologia* (FCT), Portuguese Ministry of Science and Higher Education, through a Post-doctoral fellowship (contract no. SFRH/BPD/79011/2011) and FCT project (PEst-OE/EEI/LA0009/2013.

References

[1] M. Afonso, J. Bioucas-Dias, and M. Figueiredo. Fast image recovery using variable splitting and constrained optimization. *IEEE Trans. on Im. Proc.*, 19 (9):2345–2356, 2010.

[2] A. Beck and M. Teboulle. A fast iterative shrinkage-thresholding algorithm for linear inverse problems. *SIAM Jour on Imag Sci*, 2(1):183–202, 2009.

[3] J.M. Bioucas-Dias and M. A T Figueiredo. An iterative algorithm for linear inverse problems with compound regularizers. In *Image Processing, 2008. ICIP 2008. 15th IEEE International Conference on*, pages 685–688, Oct 2008.

[4] Thomas Blumensath and Mike E Davies. Iterative thresholding for sparse approximations. *Journal of Fourier Analysis and Applications*, 14(5-6):629–654, 2008.

[5] Alan C Bovik. *Handbook of image and video processing*. Academic Press, 2010.

[6] Jian-Feng Cai, Raymond H Chan, and Mila Nikolova. Fast two-phase image deblurring under impulse noise. *Journal of Mathematical Imaging and Vision*, 36(1):46–53, 2010.

[7] A. Chambolle. An algorithm for total variation minimization and applications. *Journal of Mathematical Imaging and Vision*, 20(1):89–97, 2004.

[8] T. Chan, S. Esedoglu, F. Park, and A. Yip. Recent developments in total variation image restoration. *Handbook of Mathematical Models in Computer Vision*, pages 17–30, 2005.

[9] Tao Chen and Hong Ren Wu. Adaptive impulse detection using center-weighted median filters. *Signal Processing Letters, IEEE*, 8(1):1–3, 2001.

[10] Rafael C Gonzalez, Richard E Woods, and Steven L Eddins. *Digital image processing using MATLAB*, volume 2. Gatesmark Publishing Tennessee, 2009.

[11] J. Portilla. Image restoration through l0 analysis-based sparse optimization in tight frames. In *Image Processing (ICIP), 2009 16th IEEE International Conference on*, pages 3909–3912, Nov 2009.

[12] L. Rudin, S. Osher, and E. Fatemi. Nonlinear total variation based noise removal algorithms. *Physica D*, 60:259–268, 1992.

[13] J-L Starck, Michael Elad, and David L Donoho. Image decomposition via the combination of sparse representations and a variational approach. *Image Processing, IEEE Transactions on*, 14(10):1570–1582, 2005.

[14] Yi Wang, Arthur Szlam, and Gilad Lerman. Robust locally linear analysis with applications to image denoising and blind inpainting. *SIAM Journal on Imaging Sciences*, 6(1):526–562, 2013.

[15] Zhou Wang, Alan C Bovik, Hamid R Sheikh, and Eero P Simoncelli. Image quality assessment: From error visibility to structural similarity. *Image Processing, IEEE Transactions on*, 13(4):600–612, 2004.

[16] Ming Yan. Restoration of images corrupted by impulse noise and mixed gaussian impulse noise using blind inpainting. *SIAM Journal on Imaging Sciences*, 6(3):1227–1245, 2013.

Quis-Campi: Extending *In The Wild* Biometric Recognition to Surveillance Environments

João Neves
jcneves@penhas.di.ubi.pt
Gil Santos
gmelfe@ubi.pt

IT - Instituto de Telecomunicações, Universidade da Beira Interior

Abstract

Efforts in biometrics are being held into extending robust recognition techniques to *in the wild* scenarios. Nonetheless, and despite being a very attractive goal, human identification in the surveillance context remains an open problem. In this paper main we introduce the concept for a novel biometric system – *Quis-Campi* – that effectively bridges the gap between surveillance and biometric recognition while having a minimum amount of operation restrictions. We propose a fully automated surveillance system for human recognition purposes, attained by combining human detection and tracking, further enhanced by a PTZ camera that delivers data with enough quality to perform biometric recognition. Along with the system concept, implementation details for both hardware and software modules are provided.

1 Introduction

Biometrics is one of the most active fields in the area of computer vision, which is justified by our societies' increasing concern about security. Biometric systems significantly rely on the accurate extraction of individuals' distinctive features, which is conditioned by the acquisition environment and constraints. As such, the most reliable systems are deployed on controlled scenarios and count on subject cooperation. On the other side, surveillance cameras are widely deployed and can constitute a good source of input for biometric systems. Filling the gap between biometrics and visual surveillance is quite a desirable goal, allowing to produce *automata* capable of recognizing human beings *in the wild*, without their cooperation and, possibly, even without their awareness.

When moving to *in the wild* scenarios the acquisition constraints are substantially lowered and, most of the time, subject cooperation is not even expectable. In order to deal with such challenging conditions alternatives are sought over three axes [3]: 1) improve the existing algorithms so they can handle more degraded data; 2) resort to multi-modal biometric systems so that the usage of multiple traits can compensate for their lack of "quality"; 3) explore new biometric traits that could better cope with this new reality. Despite the recent efforts, no system yet exists capable of dealing effectively with all the issues introduced by *in the wild* biometrics, and even those systems able to cope with less constrained conditions (e.g. the Iris On The Move project [5]) still lack an ideal level of user abstraction.

Despite the several applications of visual surveillance, most of existing surveillance systems are focused on activity recognition (e.g. W^4 project [2]), and not that many of them are prepared to handle surveillance scenarios from a watchlist approach (e.g. Kamgar-Parsi *et al.* [4]). In this paper we present a novel biometric recognition system, designed to work covertly in a non-habituated and non-attended fashion, over non-standard environments. Our main goal is to conceive a system that links together both biometrics and visual surveillance, being able to conduct biometric recognition over typical surveillance scenarios, with the minimum possible amount of operation restrictions.

2 The Recognition System

The optimal recognition system would operate on any environment, thus minimizing the amount of operation restrictions. Since we are trying to bridge biometrics with the visual surveillance, we have developed our system in a typical surveillance scenario: a parking lot (Figure 3(a)). Such scenario is particularly harsh for recognition purposes, for a number of reasons: 1) it is a non-standard environment, with irregular lighting that changes not only during the day, but also accordingly to weather conditions, reflections, etc.; 2) complex background regions and the varying resolution of humans poses increasing challenges for both detection and recognition phases. 3) subjects can come from any direction, and they are rarely facing the camera which is typically placed on an upper position. Furthermore, the system should run in real-time. That being said, the recognition system we propose is devised over three main layers (Figure 1): hardware control, scene understanding, and recognition modules.

2.1 Hardware Control

To mimic a surveillance scenario, a wide-view camera was mounted on the exterior wall of a building, at a first-floor level (approximately 5m above the ground), and pointing towards a parking lot. Although this kind of camera offers a more complete overview of the scenery, it does not provide enough quality for recognition methods to work at the distances the driveway ranges from (15 to 35 meters) – Figure 2(a).

To provide recognition methods with reasonable quality data, a PTZ camera was added to the system. This way, pointing and zooming over a specific region allows acquiring a detailed view of detected subjects. As we can see from Figure 2(a), there is a substantial difference in the usable face and periocular pixel area between the two cameras as a function of the working distance. A Canon VB-H710F was used as the wide-view camera, and a Hikvision DS-2DE5286-AEL as the PTZ camera, with a framerate of 30 frames per second (fps) at maximum resolution. However, the independence between cameras demands a camera synchronization module able to map coordinates from the wide-view camera referential to the Pan-Tilt-Zoom (PTZ) coordinate system, allowing the acquisition of the high-resolution view of a portion of the scene (Figure 2(b)).

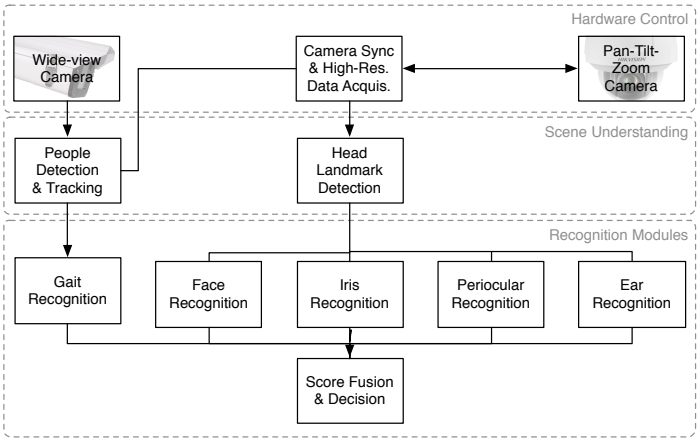
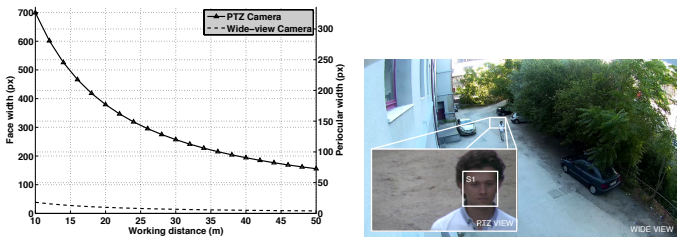


Figure 1: Working diagram of the proposed system, and the three-layer architecture: hardware control (top), scene understanding (middle) and recognition modules (bottom).



(a) Trait visibility *vs* distance (b) Illustration of the acquired data

Figure 2: Visible face and periocular width, in pixels, as function of the system's working distance (a), and illustration of the acquired data for both cameras (b).

2.2 Scene Understanding

At the scene understanding layer, the system has two main modules: 1) people detection and tracking; and 2) facial landmark detection.

The first module is responsible for locating the people as they enter the scene, and tracking them until they are no longer visible. It takes as input the video feed from the wide-view camera, and has three main steps: background subtraction, upper-body detection and tracking, illustrated in Figure 3.

Further to that, the second module of the scene understanding layer is applied to that closer view, identifying which facial landmarks are visible, thus deciding the weight of each recognition module. Being able to describe which facial traits are visible and where, is far more important that actually getting a close estimation of the head's pose, as we can tell to which extent the trait is reliable or not.

2.3 Recognition Modules

For recognition purposes the proposed system relies on a multi-modal biometric approach that combines face, iris, periocular, ear shape and gait information.

The face is not only one of the most common and widely used biometric trait, but also one of the most successful applications of image analysis and understanding. Several face recognition systems are commercially deployed, and a lot of techniques accessible [7] for both still and video images. However, as stated by Bledsoe [1], the "great variability in head rotation and tilt, lighting intensity and angle, facial expression and aging" make face recognition an extremely hard challenge. Since in surveillance scenarios the impact of such factors is even more significant, it is mandatory to rely on robust approaches (e.g. [6]).

The ocular region is one of the most explored in biometry. Iris in particular is a very popular biometric trait, delivering very high recognition accuracy under controlled environments. Although iris performance, as a biometric trait, is severely impacted in non-ideal setups, due essentially to its reduced size and moving profile, researchers are putting efforts in overcome those limitations. The periocular region represents a good trade-off between the whole face and the iris, being easy to acquire without user cooperation, and not requiring a constrained close capturing. As such, is one of the strongest candidates for the purposes of our system.

The shape of the ear can also be used as biometric trait, as the structure of its cartilage is unique for each individual. Despite all ear recognition methods traditionally require some degree of user cooperation, if proper alignment estimation can be established and the ear imaged with few or no occlusion, it can be used as biometric trait *in the wild*.

Gait is the only trait that will be imaged from the wide-view cam. Acquiring data about way a person walks is non-invasive, and can be done at-a-distance. The majority of the gait recognition methods in the literature do not require high-resolution data, so they can run over surveillance camera data.

3 Final Considerations

In this paper we present the concept of a fully automated surveillance and biometric recognition system, able to complement human detection and tracking with biometric recognition over *in the wild* surveillance environments. Although practical implementation is required, we are confident on the feasibility of such system, with special attention to its concept.

(a) Wide-view Feed (b) Background subtraction. (c) People tracking

Figure 3: Illustration of the preliminary results obtained by the people detection and tracking module: a) sample image acquired with the wide-view camera; b) foreground regions attained by background subtraction; c) people tracking module results.

3.1 Further Work

Further work should be considered over three axis: 1) a representative dataset should be acquired, not only with a large number of subjects going through typical surveillance scenes, but also with the system running over different environments (e.g. outdoor parking lot *vs* indoor lounge); 2) the software modules are still to be developed, so that the effective recognition performance can be assessed; 3) several state-of-the-art techniques should be tested for each module, and results validated over the different scenarios.

References

[1] Woodbrow W. Bledsoe. The model method in facial recognition. Technical Report PRI 15, Panoramic Research, Inc., Palo Alto, California, Palo Alto, California, 1964.

[2] I. Haritaoglu, D. Harwood, and L.S. Davis. W4: Real-time surveillance of people and their activities. *Pattern Analysis and Machine Intelligence*, 22(8):809–830, August 2000. ISSN 0162-8828. doi: 10.1109/34.868683.

[3] A.K. Jain, S. Pankanti, S. Prabhakar, L. Hong, and A. Ross. Biometrics: A grand challenge. In *Pattern Recognition (ICPR), 2004 Proc. of the 17th International Conference on*, volume 2, pages 935–942, 2004.

[4] Behrooz Kamgar-Parsi, Wallace Lawson, and Behzad Kamgar-Parsi. Toward development of a face recognition system for watchlist surveillance. *Pattern Analysis and Machine Intelligence*, 33(10): 1925–1933, October 2011. ISSN 0162-8828. doi: 10.1109/TPAMI.2011.68.

[5] J.R. Matey, O. Naroditsky, K. Hanna, R. Kolczynski, D.J. LoIacono, S. Mangru, M. Tinker, T.M. Zappia, and W.Y. Zhao. Iris on the move: Acquisition of images for iris recognition in less constrained environments. In *Proc. of the IEEE*, volume 94, pages 1936–1947, 2006. doi: 10.1109/JPROC.2006.884091.

[6] A Wagner, J. Wright, A Ganesh, Zihan Zhou, H. Mobahi, and Yi Ma. Toward a practical face recognition system: Robust alignment and illumination by sparse representation. *Pattern Analysis and Machine Intelligence*, 34(2):372–386, February 2012. ISSN 0162-8828. doi: 10.1109/TPAMI.2011.112.

[7] W. Zhao, R. Chellappa, P. Phillips, and A. Rosenfeld. Face recognition: A literature survey. *ACM Computing Surveys*, 35(4):399–458, 2000.

Partial Face Recognition Under Occlusion Using Universal Background Models

João C. Monteiro
joao.carlos.monteiro@fe.up.pt
Jaime S. Cardoso
jaime.cardoso@inescporto.pt
INESC TEC

Abstract

The rising challenges in the field of face recognition are leading to a renewed interest in the area. The present work proposes an alternative to face recognition under the effect of occlusions. An automatic modeling of SIFT descriptors, regardless of the number of detected keypoints and using a GMM-based Universal Background Model method, is proposed. Such approach allows a tight coupling between individual models and a robust likelihood-ratio decision step. The algorithm was tested on the AR databases and presented state-of-the-art performance.

1 Introduction

Over the past few years, face recognition has been on the spotlight of many research works in biometrics. The *face* is an easily acquirable trait with a high degree of uniqueness. These marked advantages, however, fall short when low-quality images are presented to the system.

It has been noted that the performance of face recognition algorithms is severely compromised when dealing with non-ideal scenarios such as non-uniform illumination, pose variations, occlusions, expression changes and radical appearance changes [3]. Whereas technological improvements in image capturing and transmitting equipment managed to attenuate most noise factors, partial face occlusions still pose a genuine challenge to automated face recognition [1].

In the present work we propose a robust alternative to face recognition under partial occlusions. Our proposal is based on the idea of maximum *a posteriori* adaptation of Universal Background Model, as proposed by Reynolds *et al.* for speaker verification [5]. We evaluate the proposed algorithm on a dataset characterized by a variety of occlusions and facial expressions.

2 Proposed Methodology

2.1 Algorithm Overview

The proposed algorithm is schematically represented in Figure 1. During enrollment a new individual's biometric data is inserted into a previously existent database of individuals. Such database is probed during the recognition process to assess either the validity of an identity claim - *verification* - or the k most probable identities - *identification* - given an unknown sample of biometric data.

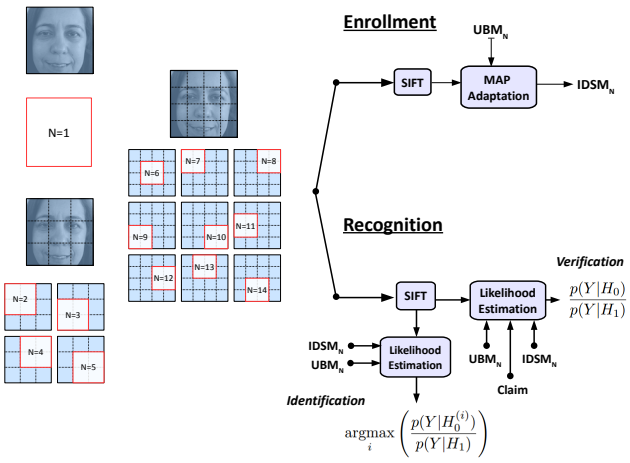


Figure 1: Example images from one subject of the AR database.

A set of M partial face models is trained for each subject, as exemplified in Figure 1. By training models that only describe features of limited regions of the face we confer our algorithm a robustness to events where all other regions are occluded. The full database of N individuals is, thus, composed by a total of $M \times N$ individual models, obtained by maximum *a posteriori* (MAP) adaptation of a corresponding set of M Universal Background Models (UBM). Both UBM and MAP adaptation will be further motivated in the following sections.

The recognition phase is carried out using the projection of the features extracted from an unknown sample onto both the M UBM and the corresponding M individual specific models (IDSM) of interest. A likelihood-ratio between both projections yields a set of M recognition scores, that are then used to perform a joint decision.

2.2 Universal Background Model

The motivation behind the UBM strategy can be easily understood if the problem of biometric verification is interpreted as a basic hypothesis test. Given a biometric sample Y and a claimed ID, S, we define:

$$H_0\text{: } Y \text{ belongs to } S$$
$$H_1\text{: } Y \text{ does } \underline{\text{not}} \text{ belong to } S$$

as the null and alternative hypothesis, respectively. The optimal decision is taken by a *likelihood-ratio test*:

$$\frac{p(Y|H_0)}{p(Y|H_1)} \begin{cases} \geq \theta & \text{accept } H_0 \\ \leq \theta & \text{accept } H_1 \end{cases} \tag{1}$$

where θ is the decision threshold for accepting or rejecting H_0, and $p(Y|H_i)$, $i \in \{0,1\}$ is the likelihood of observing sample Y when we consider hypothesis i to be true.

Biometric recognition can, thus, be reduced to the problem of computing the likelihood values $p(Y|H_0)$ and $p(Y|H_1)$. It is intuitive to note that H_0 should be represented by a model λ_{hyp} that characterizes the hypothesized individual, while, alternatively, the representation of H_1, $\lambda_{\overline{hyp}}$, should be able to model *all the alternatives to the hypothesized individual*.

From such formulation rises the need for a model that successfully covers the space of alternatives to the hypothesized identity. The most common designation in literature for such a model is *universal background model* or *UBM* [6]. Such model must be trained on a large set of data, so as to faithfully cover a representative user space and a significant amount of sources of variability. Individual models, λ_{hyp}, specialized to each enrolled subject, can then be adapted from the UBM in a fast and robust way by MAP adaptation.

2.3 Hypothesis Modeling

On the present work we chose Gaussian Mixture Models (GMM) to model both the UBM, i.e. $\lambda_{\overline{hyp}}$, and the individual specific models (IDSM), i.e. λ_{hyp}. This choice is extensively motivated in many works of related areas. From the most common interpretations, GMMs are seen as capable of representing broad "hidden" classes, reflective of the unique structural arrangements observed in the analyzed biometric traits [5]. Besides this assumption, Gaussian mixtures display both the robustness of parametric unimodal Gaussian density estimates, as well as the ability of non-parametric models to fit non-Gaussian data [4].

All models are trained on sets of Scale Invariant Feature Transform (SIFT) keypoint descriptors. The invariance of SIFT features to many undesirable factors such as scale, translation, rotation and partially to illumination, affine and 3D transformation, confer them a strong appeal in the area of unconstrained biometrics.

2.3.1 MAP Adaptation

IDSM are generated by the *tuning of the UBM parameters*, in a maximum *a posteriori* (MAP) sense, using individual specific biometric data. This approach provides a tight coupling between the IDSM and the UBM, resulting in better performance and faster scoring than uncoupled methods, as well as a robust and precise parameter estimation, even when only a small amount of data is available [7]. The adaptation process consists in two main estimation steps. First, for each mixture of the UBM, a set of sufficient statistics are computed from a set of M individual specific feature vectors, $X = \{\mathbf{x}_1 ... \mathbf{x}_M\}$:

$$n_i = \sum_{m=1}^{M} p(i|\mathbf{x}_m) \tag{2}$$

$$E_i(\mathbf{x}) = \frac{1}{n_i} \sum_{m=1}^{M} p(i|\mathbf{x}_m)\mathbf{x}_m \tag{3}$$

$$E_i(\mathbf{x}\mathbf{x}^t) = \frac{1}{n_i} \sum_{m=1}^{M} p(i|\mathbf{x}_m)\mathbf{x}_m\mathbf{x}_m^t \tag{4}$$

where $p(i|\mathbf{x}_m)$ represents the probabilistic alignment of $\mathbf{x}_m$ into each UBM mixture. Each UBM mixture is then adapted using the newly computed sufficient statistics, and considering diagonal covariance matrices. The update process can be formally expressed as:

$$\hat{w}_i = [\alpha_i n_i/M + (1-\alpha_i)w_i]\,\xi \tag{5}$$

$$\hat{\mu}_i = \alpha_i E_i(\mathbf{x}) + (1-\alpha_i)\mu_i \tag{6}$$

$$\hat{\Sigma}_i = \alpha_i E_i(\mathbf{x}\mathbf{x}^t) + (1-\alpha_i)(\sigma_i\sigma_i^t + \mu_i\mu_i^t) - \hat{\mu}_i\hat{\mu}_i^t \tag{7}$$

$$\sigma_i = \mathrm{diag}(\Sigma_i) \tag{8}$$

where $\{w_i, \mu_i, \sigma_i\}$ are the original UBM parameters and $\{\hat{w}_i, \hat{\mu}_i, \hat{\sigma}_i\}$ represent their adaptation to a specific speaker. To assure that $\sum_i w_i = 1$ a weighting parameter ξ is introduced. The α parameter is a data-dependent adaptation coefficient. Formally it can be defined as $\alpha_i = \frac{n_i}{r+n_i}$. The relevance factor r weights the relative importance of the original values and the new sufficient statistics. In the present work we set $r = 16$.

2.4 Joint Decision

The whole training process is repeated M times for each of the M partial face regions, defined *a priori*. In the present work we used $M = 14$ as exemplified in Figure 1. During the recognition step the unknown test image is cropped into the M alternative partial face regions and each recognition score against the corresponding UBM and IDSM is carried out. This yields a total of M likelihood-ratio scores for each test image and hypothesized identity. The joint decision is carried out by using the *sum* of the local likelihood-ratio scores.

3 Results and Discussion

3.1 Tested Dataset and Experimental Setup

We chose to assess the performance of the proposed algorithm of the AR database [2]. This database contains over 4000 frontal face images from 126 individuals, acquired under variable illumination, expression, and occlusion. Occlusions can be divided into two main categories: *sunglasses* and *scarf*. An example of all images from a single individual is presented in Figure 2.

Figure 2: Example images from one subject of the AR database.

All unoccluded images from every individual are chosen to train both the local UBMs and IDSMs. The remaining scarf and sunglass occluded images are tested separately, so as to better allow an analysis of the consistency of the proposed algorithm when exposed to variable types of occlusion.

3.2 Recognition Performance

The most relevant results regarding the biometric performance of the proposed algorithm are summarized in the first line of Table 1. The chosen metrics were the rank-1 recognition rate (R_1) and the equal error rate (EER). The R_1 refers to the rate of correctly identified individuals, whereas the EER is observed at the decision threshold, θ, where the errors of falsely accepting and falsely rejecting H_0 occur with equal frequency.

Work	Sunglasses		Scarf		Average	
Proposed	95.67%	0.0327	**96.50**%	0.0251	**96.09**%	0.0289
Min *et al.* [3]	75.00%	–	92.08%	–	83.54%	–
Li *et al.* [1]	**97.00**%	–	52.00%	–	74.50%	–
	R_1	EER	R_1	EER	R_1	EER

Table 1: Comparison between the obtained results for the AR database and some state-of-the-art algorithms.

Table 1 also presents a few results from some recent state-of-the-art works on the AR database. It is readily observed that the proposed algorithm presents the most consistent and robust behavior, regardless of the nature of the present occlusions. While the work by Li *et al*. presents higher performance for sunglasses occlusion, their performance regarding the alternative scarf occlusion is considerably lower than the one obtained with the proposed algorithm. Min *et al*. present a work that seems to have exactly the opposite problem, with good performance observed for scarf occlusion, but lower results for sunglasses. The proposed algorithm, thus, presents the significant advantage of being able to achieve high performance regardless of the nature and location of the occlusion by performing recognition on a set of robust local descriptors.

4 Conclusion and Future Work

In the present work we propose an alternative to face recognition when the images present occluded regions, thus hindering the functioning mode of the most traditional methodologies. The obtained results in the AR database indicate the robust and consistent nature of our algorithm, independent of the nature and location of the occlusions. One effect that we have not considered is how variable illumination might cause a degradation in performance. As the chosen SIFT descriptors are known to be only partially invariant to illumination, high variations of such parameter might lead to the aforementioned nefarious effect. Furthermore, as only a few locations of the image contain meaningful information for the biometric recognition step it would be interesting to find a more robust and efficient alternative to the simple sum-rule joint decision presented in this work.

Acknowledgment

This work is financed by the ERDF European Regional Development Fund through the COMPETE Programme (operational programme for competitiveness), FCOMP-01-0124-FEDER-037281 and by National Funds through the FCT Fundação para a Ciência e Tecnologia (Portuguese Foundation for Science and Technology) within project PEst-C/EEI/LA0014/2013 and PhD grant SFRH/BD/87392/2012.

References

[1] Yue Long Li, Li Meng, Ju Fu Feng, and Ji Gang Wu. Downsampling sparse representation and discriminant information aided occluded face recognition. *Science China Information Sciences*, 57(3):1–8, 2014.

[2] Aleix M Martinez. The AR face database. *CVC Technical Report*, 24, 1998.

[3] Rui Min, Abdenour Hadid, and Jean-Luc Dugelay. Efficient detection of occlusion prior to robust face recognition. *The Scientific World Journal*, 2014, 2014.

[4] D. Reynolds. Gaussian mixture models. *Encyclopedia of Biometric Recognition*, pages 12–17, 2008.

[5] D.A. Reynolds, T.F. Quatieri, and R.B. Dunn. Speaker verification using adapted gaussian mixture models. *Digital signal processing*, 10(1):19–41, 2000.

[6] Douglas A Reynolds. An overview of automatic speaker recognition technology. In *Acoustics, Speech, and Signal Processing (ICASSP), 2002 IEEE International Conference on*, volume 4, pages IV–4072. IEEE, 2002.

[7] K. Shinoda and N. Inoue. Reusing speech techniques for video semantic indexing [applications corner]. *Signal Processing Magazine, IEEE*, 30(2):118–122, 2013. ISSN 1053-5888. doi: 10.1109/MSP.2012.2230520.

Liveness detection methods in fingerprint recognition

Ana F. Sequeira
afps@inescporto.pt
Ana Rute Louro
ee08206@fe.up.pt
Jaime S. Cardoso
jaime.cardoso@inescporto.pt

INESC TEC and Faculdade de Engenharia,
Universidade do Porto,
Porto, Portugal

Abstract

Biometric recognition is nowadays a mature technology with several applications. However, biometric systems based on fingerprint are vulnerable to direct attacks consisting on the presentation of a fake fingerprint to the sensor. This work focuses on fingerprint liveness detection methods as an attempt to overcome that vulnerability. Two methods from the state-of-the-art in iris liveness detection were tested with fingerprint databases containing different kinds of fake samples. One aim of the work was to investigate how these iris techniques would perform with fingerprint fake samples. The other purpose was to diversify the classification scenario by broaden the classification task from being made within each type of samples to being made in sets mixing the types of fake samples.

1 Introduction

Biometric recognition is a mature technology used in many government and civilian applications such as e-passports, ID cards, and border control. Examples include the US-Visit (United States Visitor and Immigrant Status Indicator Technology) fingerprint system, the Privium iris system at Schiphol airport, and the SmartGate face system at Sydney Airport [1]. Fingerprints recognition systems (FRS) are widely used, in fact, probably this was the first biometric trait to be used to identify people. Fingerprints are small lines/ridges and valleys in the skin of fingertips. Their configuration is formed at around the seventh month of fetus development due to a combination of genes and environmental factors and do not change throughout life (except if an accident, such as a severe burnt, happens) [5, 6]. The influence of environmental factors in the fingerprint formation results in such variations in its configuration that it is considered impossible to have two fingerprint looking exactly alike [5, 9].

Some fingerprint recognition techniques use correlation-based methods to directly compare images but most of the fingerprint recognition and classification algorithms employ a feature extraction stage [5]. Also, some preprocessing, segmentation and enhancement steps are often performed to simplify the task of minutiae extraction.The main steps of a FRS system may bee seen in Fig. 1.

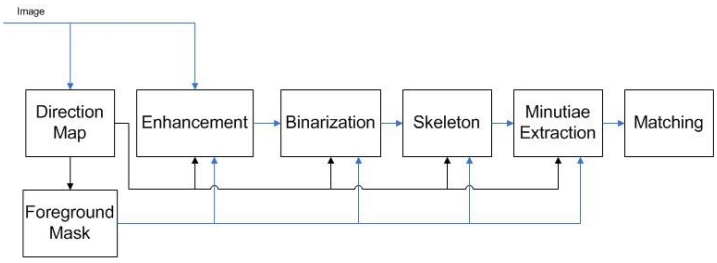


Figure 1: Fingerprint recognition system block diagram.

Biometric recognition systems in general, and FRS in particular, can be spoofed by presenting fake samples of the biometric trait to the sensor used in a specific system. The fake samples can be acquired with or without user cooperation: an authorized user may help an hacker to create a clone of his fingerprint; or the fingerprint is obtained from a glass or other surface (latent fingerprints) [5, 7]. Latent fingerprints can be painted with a dye or powder and then "lifted" with tape or glue. However, these prints are, usually, low quality as they can be incomplete or smudged and thus are not very accurate. The easiest way of creating a fake sample is by printing the fingerprint image into a transparent paper. A more successful method is to create a 3D fake model with the fingerprint stamped on it. This can be done by creating a mold that is then filled with a substance (silicon, gelatin, Play-Doh, wax, glue, plastic). Then this mold is used to create a thick or thin dummy that an intruder can use, an example is depicted in Fig. 2.

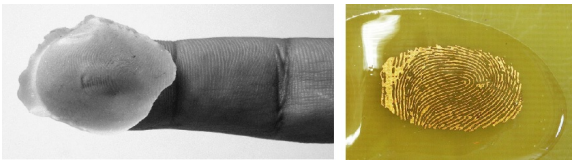


Figure 2: Finger model and mold (Adapted from [5, 6]).

Texture features are good descriptors for fingerprints, as they are for iris. Therefore, some liveness detection methods for both these traits are texture-based. However, detecting the liveness of a fingerprint is even harder than detecting the liveness of an iris, as both the material used in the fake models and the epidermis of a finger are not alive. Liveness detection can be performed through additional hardware or by processing the obtained image. Hardware-based solutions work by measuring some physical characteristics (such as blood pressure, temperature, pulse, or pupil dilation, voluntary eye blink, among others) and have the disadvantage of being expensive and requiring intervention at the device level. We confined this work to software-based approaches. In this work we applied two methods previously used in iris liveness detection in fingerprints images with fake samples to access their performance in fingerprint liveness detection. Having in mind that in real-world solutions the system are not "aware" of the kind of attack that might be performed, we chose to merge fake samples forged with different materials to compose a new set of images captured by the same sensor but comprising different materials for the fake samples.

2 Implemented Methods

2.1 Weighted Local Binary Patterns (wLBP)

The Weighted Local Binary Patterns (wLBP) method [10] combines Local Binary Patterns (LBP) with a Scale Invariant Feature Transform (SIFT) [4]. The process starts with the generation of a Gaussian scale space. The output of this operation is a smothed image in six scales. For each scale, the gradient orientation of each pixel is calculated and an histogram of gradient orientations is created. Every histogram is then converted into a descending rank, from 7 to 0. The LBP method labels the pixels of an image by comparing them with their neighborhood. Combining the LBP method with SIFT results in weighted maps. Three simple maps are constructed using the 3 first scales and a fourth map results from the combination of the last scales. Each map is divided in 8 by 8 blocks, the first and last lines of blocks are discarded in order to avoid obstructions, and three statistical features are extracted from each block (standard deviation of the wLBP histogram, mean of wLBP map and standard deviation of the wLBP map). That results in a 576 dimensional feature.

2.2 Gray Level Co-occurence Matrices (GLCM)

The Gray Level Co-occurence Matrices (GLCM) method [8] is based on Gray Level Co-occurence Matrices (GLCM) which characterize the relationship between neighboring pixels. Fourteen features are extracted from each GLCM matrix [3]: angular second moment, contrast, correlation, variance, inverse difference moment, sum average, sum variance, sum entropy, entropy, difference variance, difference entropy, two information measures of correlation and the maximal correlation coefficient. These features are orientation dependent, so, four different values can be

	Biometrika									
	Ecoflex		Gelatin		Latex		Modasil		Wood Glue	
	μ	σ	μ	σ	μ	σ	μ	σ	μ	σ
wLBP	**0.78**	0.22	3.82	0.61	1.54	0.40	0.98	0.38	1.38	0.38
GLCM	**0.92**	0.31	2.29	0.58	3.44	0.544	3.66	0.91	4.82	0.71

Table 1: Results of method 1, for Biometrika (errors in %).

obtained corresponding to each feature based on each of the following four orientations (0°, 45°, 90°, 135°). The mean and standard deviation of the four values (four orientations) of each 14 measures, compose a set of 28 features.

3 Experimental Setup

3.1 Datasets

The methods were tested in two subsets of the fingerprint datasets made available for the LivDet2013 competition. The images of these datasets were acquired with four different sensors: Biometrika, Crossmatch, Italdata and Swipe and the fake samples were built using seven different materials: Body Double, Latex, Play-Doh, Wood Glue, Gelatin, Silicon and Modasil. For more details on these sets see [2]. The subsets tested for this work were **Biometrika** and **Swipe**. This choice was made based on previous results in which Biometrika and Swipe performed better for wLBP and GLCM methods, respectively. Biometrika subset comprise 2000 real samples and 400 samples for each material. For building the fake samples of Biometrika the materials used were: Ecoflex, Gelatine, Latex, Modasil and Wood Glue. Swipe subset comprise 2374 real samples and approximately 500 samples for each material. For the fake samples of Swipe the materials used were: Latex, Wood Glue, Body Double and Play-Doh.

3.2 Methodology

The **feature extraction** was performed using the two methods, wLBP and GLCM (see section 2) in the whole image. For the **classification** task we used Support Vector Machines (SVM), with a polynomial kernel, and for optimizing the parameters a "grid-search" was performed on C and d parameters: the exponential growth of $C = 2^N$ was tested, with N varying from -1 to 15 and the polynomial degree (d) was tested with the following values $\{1,2,3,4,5\}$. We also used cross-validation so that the data was divided randomly in 62.5% of the samples for training and 37.5% for testing. Two different classification scenarios were studied: within each material (Method 1) and using a mix of all materials (Method 2 -"mixed sets"). In methods 1 and 2, the classification error was obtained by averaging the error in 50 runs and in each run the data was divided randomly in 62.5% of the samples for training and 37.5% for testing.

3.3 Experimental Results and Discussion

In the first classification scenario, which is the typical one in literature and in most competitions in the field, the Biometrika and the Swipe subsets are evaluated testing the different materials separately (Method 1). The results obtained are presented in Table 1 and Table 2, respectively.

	Swipe							
	Body Double		Latex		Play-Doh		Wood Glue	
	μ	σ	μ	σ	μ	σ	μ	σ
wLBP	6.93	0.78	8.78	0.93	5.51	0.92	**5.23**	0.68
GLCM	4.53	0.64	3.60	0.67	3.54	0.74	**3.51**	0.77

Table 2: Results of method 1, for Swipe (errors in %).

Observing the results of Table 1, we may observe that, for Biometrika, the best results of the both methods are obtained using the same material, Ecoflex, 0.78% and 0.92%, respectively. In Table 2, for Swipe, we may observe that the best results of both methods are also obtained using the same material, Wood Glue, 5.23% and 3.51%, respectively. In an overall analysis, better results are obtained with the Biometrika subset and, in general, the wLBP method leads to better results.

Considering that, in practice, the system is not aware a priori of the kind of fake sample to be used in an attempt to fool the system the classification performed in method 1 is not very realistic. Method 2, presented in section 3.2, was applied aiming to overcome this limitation. In this scenario, the training and test sets comprise all the known varieties of different materials used to build the fake samples. On one hand, we gather the whole information but, on the other hand, we do not train our system to face a single material. We do not restrict the study to one material at a time, as traditional approaches do. The results obtained for the Biometrika and the Swipe subsets using method 2 are presented in Table 3.

	Biometrika		Swipe	
	μ	σ	μ	σ
wLBP	2.71	0.35	8.37	0.60
GLCM	5.00	0.62	6.61	0.66

Table 3: Results of method 2 for Biometrika and Swipe datasets (errors in %).

Comparing the results of Tables 1 and 2 with the results of Table 3, we may observe that, for both methods, WLBP and GLCM, the best result is given by method 1, in both Biometrika and Swipe datasets. Therefore we conclude that performing the classification task within each material separately leads to better results than using a set with all the fake samples mixed together.

4 CONCLUSIONS AND FUTURE WORK

In this work, two methods for iris liveness detection were applied in fingerprint images. Comparing our results with the ones from LivDet2013[2] we may consider our results encouraging of further investigation since in some cases the results outperform those. Regarding the two different classification scenarios, we concluded that the results worsened when we go from training and testing within the same fake samples to mixing all the materials (but fixing the sensor). This is not unexpected since the variability in the types of fake samples is expected to increase the difficulty of the classification task. This finding leads us to think that the traditional approach we find in the literature is a somehow optimist. As future work we intend to broad this study to more datasets for fingerprint liveness detection and possibly compare this methods with state-of-the-art methods in this field.

Acknowledgements

This work is financed by the ERDF European Regional Development Fund through the COMPETE Programme (operational programme for competitiveness), FCOMP-01-0124-FEDER-037281 and by National Funds through the FCT - Portuguese Foundation for Science and Technology - within project PEst-C/EEI/LA0014/2013 and the first author's PhD grant SFRH/BD/74263/2010.

References

[1] Fernando Alonso-Fernandez, Julian Fierrez, and Javier Ortega-Garcia. Quality measures in biometric systems. *IEEE Security & Privacy*, 10(6):52–62, 2012.

[2] Luca Ghiani, David Yambay, Valerio Mura, Simona Tocco, Gian Luca Marcialis, Fabio Roli, and Stephanie Schuckers. Livdet 2013 fingerprint liveness detection competition. In *Int. Conf. on Biometrics (ICB)*, pages 1–6, 2013.

[3] R.M. Haralick, K. Shanmugam, and I. Dinstein. Textural features for image classification. In *IEEE TSMC*, volume SMC-3, no. 6, pages 610 – 621, 1973.

[4] David G Lowe. Object recognition from local scale-invariant features. In *7th IEEE Int. Conf. on Comp. Vis.*, volume 2, pages 1150–1157, 1999.

[5] Davide Maltoni, Dario Maio, Anil K Jain, and Salil Prabhakar. *Handbook of fingerprint recognition*. Springer, 2009.

[6] Marie Sandström. Liveness detection in fingerprint recognition systems. 2004.

[7] Ton Van der Putte and Jeroen Keuning. Biometrical fingerprint recognition: don't get your fingers burned. In *Smart Card Research and Advanced Applications*, pages 289–303. Springer, 2000.

[8] Zhuoshi Wei, Xianchao Qiu, Zhenan Sun, and Tieniu Tan. Counterfeit iris detection based on texture analysis. In *ICPR 2008. 19th International Conference on Pattern Recognition.*, pages 1–4. IEEE, 2008.

[9] Yau Wei Yun. The "123" of biometric technology. *Synthesis Journal*, 2002.

[10] Hui Zhang, Zhenan Sun, and Tieniu Tan. Contact lens detection based on weighted LBP. In *20th Int. Conf. on Patt. Rec. (ICPR)*, pages 4279–4282. IEEE, 2010.

Detecção de Veículos e Edifícios

Jorge Leitão[1,2]
08152.leitao@academiamilitar.pt
José Bioucas-Dias[2]
bioucas@lx.it.pt
José S. Silva[1]
jose.silva@academiamilitar.pt
Miguel Gonçalves[1,2]
08218.goncalves@academiamilitar.pt

[1] Departamento de Ciências e Tecnologias de Engenharia
Academia Militar
Lisboa, Portugal

[2]Departamento de Engenharia Electrotécnica e de Computadores
Instituto Superior Técnico
Lisboa, Portugal

Resumo

A detecção de veículos e edifícios através de imagens aéreas é um problema complexo em visão computacional e detecção remota. A maioria dos trabalhos de investigação nesta área aborda apenas um destes temas. Neste trabalho é proposta uma abordagem para resolver ambos os problemas, baseada na transformada de Hough, características SIFT e classificadores MSV (Máquinas de Suporte Vectorial), usando imagens aéreas de alta resolução obtidas a partir de um UAV (*Unmanned Aerial Vehicle*). Duas componentes importantes da metodologia desenvolvida são a identificação de zonas com veículos e a identificação de telhados rectangulares com transformada de Hough. Os resultados experimentais mostram que o algoritmo proposto é robusto em ambos os problemas.

1 Introdução

A detecção de veículos e de edifícios através de imagens aéreas é um problema actual em visão computacional e em aplicações de detecção remota. A investigação nesta área tem aumentado significativamente devido à sua aplicação em sistemas geoespaciais, planeamento e monitorização urbana, cartografia e aplicações militares, nomeadamente seguimento e localização de infraestruturas sensíveis, entre outros [1] e [2]. Têm sido propostas várias metodologias, desde as mais simples até às mais complexas como a fusão de diferentes métodos. Na detecção de veículos é comum a utilização de classificadores. Em [3], após a identificação de zonas potencialmente possuidoras de veículos através do *Harris corner detector*, são extraídas características baseadas em Histogramas de Coeficientes de Gabor, e de seguida é usado um classificador MSV ou *K-Nearest Neighbors*. Com base nos algoritmos e resultados obtidos em [3], Moranduzzo [4] propôs outra abordagem mas com resultados melhores, em que as características de cor e SIFT são recolhidas e posteriormente classificadas com Máquinas de Suporte Vectorial. Esta metodologia, recorrendo a classificadores, é também aplicável à detecção de edifícios. Em [5] é proposta a realização de uma segmentação inicial com o algoritmo *mean-shift*, seguida de extracção de características de cor e *multi-scale orientation histogram* e posteriormente a classificação com Máquinas de Suporte Vectorial. Uma metodologia que não recorre a classificadores, que diminui o tempo de computação e exclui o uso de conjuntos de treino, é proposta por Pakizeh [6]. Inicialmente os centros dos edifícios são estimados através da intensidade da imagem, confirmando-se ou não a sua existência através da forma, recorrendo à transformada de Hough.

Neste trabalho propõe-se uma abordagem simultânea a ambos os problemas, tal como [1], baseando-se na Transformada de Hough e em características SIFT para a detecção e identificação de Edifícios e Veículos, respectivamente. Esta abordagem efectua a detecção de edifícios sem recurso a classificadores e apresenta uma abordagem alternativa que identifica, com elevada probabilidade, os locais onde existem veículos.

2 Metodologia Proposta

As imagens RGB utilizadas foram obtidas por UAV voando a baixa altitude sobre a Academia Militar na Amadora e produzindo imagens com alta resolução (3000x4000 pixéis).

A metodologia proposta encontra-se indicada na Figura 1. Dividindo o problema, a imagem que se pretende discriminar é utilizada em duas partes distintas: detecção de veículos e detecção de edifícios. Na primeira, são extraídas características SIFT e informação de cor que são classificadas através de máquinas de suporte vectorial. Aos resultados da classificação é aplicado um algoritmo de fusão de pontos importantes com intuito de obter a localização mais provável de veículos. Na segunda parte da metodologia proposta, após um pré-processamento baseado na cor dos telhados dos edifícios, usa-se a transformada de Hough para a detecção de rectas e um algoritmo para detecção de formas rectangulares.

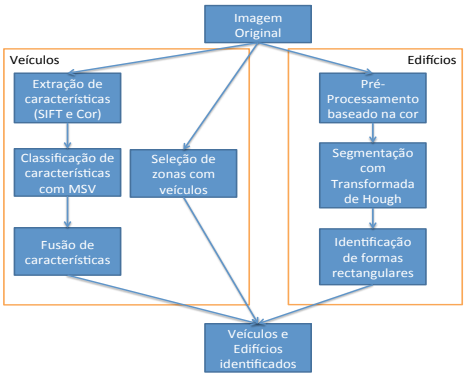


Figura 1: Metodologia Proposta

2.1Veículos

A) EXTRACÇÃO DE CARACTERÍSTICAS (SIFT e COR)

Para a extracção de características SIFT usou-se a biblioteca *open source* VLFeat que, entre outros, implementa o algoritmo SIFT, descrito em [7]. Após este procedimento, é extraída informação de cor (RGB e HSV) dos pixéis identificados como pontos de interesse pelo SIFT, sendo construído um vector de 134 elementos: SIFT+RGB+HSV (128 + 3 + 3). São considerados pontos de interesse pelo algoritmo SIFT aqueles cujo vector de características seja invariante a rotação, mudanças de intensidade e de escala, e transformações afim, numa imagem.

B) CLASSIFICAÇÃO DE CARACTERÍSTICAS COM MSV

Constituído o conjunto de pontos importantes da imagem, é usada uma máquina de suporte vectorial com o objectivo de identificar os pontos importantes que pertencem a veículos. Admite-se a existência de um veículo quando a percentagem de pontos importantes classificados como tal é superior a um determinado limiar definido pelo utilizador. O veículo é identificado como o ponto médio dos pontos importantes pertencentes à classe veículo.

C) FUSÃO DE CARACTERÍSTICAS

Conhecendo todos os pontos com elevada probabilidade de existência de veículos, e uma vez que estes são em maior número do que a quantidade de veículos existentes, é aplicado um algoritmo para os agrupar por veículo. O critério de agrupamento, inspirado em [4], baseia-se na distância entre pontos. É definido um limiar de distância para serem considerados pontos vizinhos e sempre que a distância seja menor que esse limiar, dois pontos são fundidos num só. O processo repete-se enquanto existirem pontos menos distantes entre si que o limiar definido.

D) SELECÇÃO DE ZONAS COM VEÍCULOS

Nesta fase, são identificadas zonas da imagem com grande probabilidade de existência de veículos. Para tal, considera-se que o tejadilho de um veículo tem, tipicamente, forma quadrangular e que os

seu vidros, observados por vista aérea, são escuros. Assim, admite-se que zonas com potenciais veículos, são constituídas por dois quadrados (interno e externo). Estas zonas são identificadas na imagem se o desvio padrão da intensidade dos pixéis contidos em cada quadrado for pequeno e se os pixéis contidos no quadrado externo forem, maioritariamente escuros. Por fim, são identificados os veículos resultantes da fase C que pertençam às zonas selecionadas.

2.2 Edifícios

A) PRÉ-PROCESSAMENTO

O pré-processamento é realizado com base na cor, neste caso, vermelho, cor predominante nos telhados. Utilizando a imagem com componentes R, G e B, a segmentação baseia-se no facto de o vermelho se caracterizar por ter a componente R elevada comparativamente à G e B num determinado pixel. A segmentação é dada por:

$$vermelho = R - \max(G, B) \qquad (1)$$

B) SEGMENTAÇÃO COM TRANSFORMADA DE HOUGH

À imagem obtida na fase anterior, é aplicado o detector de arestas de Canny. De seguida a imagem é percorrida por uma janela quadrangular sobre a qual é aplicada a transformada de Hough e posteriormente procuradas formas rectangulares.

O detector de Canny é um método que permite a detecção de arestas numa imagem com base na direcção e intensidade do gradiente calculado em todos os pixéis.

A transformada de Hough é um método para detecção de formas parametrizáveis (círculos, linhas, elipses...). A transformada é aplicada a todos os pontos de uma imagem, de tal modo que os pertencentes à forma geométrica a parametrizar sejam mapeados no mesmo ponto do espaço de parametrização.

C) IDENTIFICAÇÃO DE FORMAS RECTANGULARES

A identificação de formas rectangulares é realizada com base nos segmentos de recta detectados com a transformada de Hough. O algoritmo desenvolvido detecta um potencial edifício com forma rectangular, se existirem quatro segmentos de recta $\overline{AB}, \overline{CD}, \overline{EF}, \overline{GH}$ que verifiquem todos os requisitos, condicionados por limiares:

1. $\overline{AB} \parallel \overline{CD} \wedge \overline{EF} \parallel \overline{GH}$
2. $\|AB\| \cong \|CD\| \wedge \|EF\| \cong \|GH\|$
3. $\overline{AB} \perp \overline{EF}$
4. $car \begin{bmatrix} \|AE\| & \|AF\| & \|AG\| & \|AH\| \\ \|BE\| & \|BF\| & \|BG\| & \|BH\| \\ \|CE\| & \|CF\| & \|CG\| & \|CH\| \\ \|DE\| & \|DF\| & \|DG\| & \|DH\| \end{bmatrix} = 4$

Nota: $\overline{AB}$, segmento de recta; $\parallel$, paralelo a; $\wedge$, conjunção; $\cong$, aproximadamente igual a; $\|AB\|$, comprimento do segmento de recta $\overline{AB}$; $\perp$, perpendicular a; $car[A]$, característica da matriz A (número máximo de linhas/colunas linearmente independentes).

3 Resultados

Na detecção de edifícios, é factor condicionante o limiar que define o nível de vermelho a segmentar. O limiar definido condiciona a forma do telhado que será usada pelo detector de arestas e consequentemente a detecção de formas rectangulares. A segmentação baseada na cor não tem a robustez ideal uma vez que o tom de vermelho varia entre telhados. A transformada de Hough, apesar de depender de parâmetros ajustáveis que condicionam os resultados, é robusta para arestas bem definidas e, neste caso, para formas rectangulares, como se verifica na Figura 2.

No que respeita à detecção de veículos verifica-se, para a Figura 2, uma taxa de detecção (TD) de 75% calculada a partir de TD=VP/(VP+VN), onde VP (verdadeiros positivos) expressa o número de veículos detectados e VN (verdadeiros negativos) corresponde aos veículos presentes na imagem que não foram identificados pelo algoritmo. Estes resultados são justificáveis com a MSV binária utilizada uma vez que é complexo treinar o classificador com todas as classes passíveis de existir numa imagem.

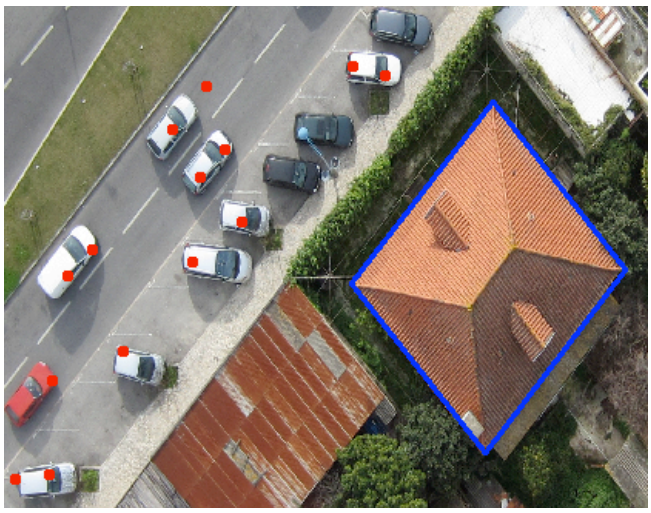


Figura 2: Veículos (pontos vermelhos) e Edifício (delimitado a azul) detectados

4 Conclusões

Neste trabalho foi desenvolvido uma metodologia para detecção de edifícios e veículos em imagens aéreas. Para a detecção de veículos é usado um classificador baseado numa máquina de suporte vectorial, seguida de um algoritmo de fusão por forma a obter um ponto por veículo. Relativamente a edifícios, utilizou-se a transformada de Hough após pré-processamento baseado na cor.

O algoritmo proposto apresenta uma abordagem simples e eficiente na detecção de rectângulos através da transformada de Hough e uma nova forma de seleção de zonas com veículos, constituindo-se uma ferramenta robusta na resolução de ambos os problemas de detecção. Com a metodologia proposta, são identificados 9 em 12 veículos, bem como o edifício presente na imagem.

Agradecimentos

Ao CINAMIL, Centro de Investigação da Academia Militar, através do projeto FUSIMIL, pelo financiamento que suportou as despesas relativas à conferencia RECPAD'2014.

Referências

[1] J. Porway, Q. Wang, and S. Zhu, "A hierarchical and contextual model for aerial image Parsing," *International Journal of Computer Vision,* vol. 88, pp. 254-283, 2010.

[2] M. Cote and P. Saeedi, "Automatic rooftop extraction in nadir aerial imagery of suburban regions using corners and variational level set evolution," *IEEE Transactions on Geoscience and Remote Sensing,* vol. 51, pp. 313-328, 2013.

[3] J. Gleason, A. V. Nefian, X. Bouyssounousse, T. Fong, and G. Bebis, "Vehicle detection from aerial imagery," in *IEEE International Conference on Robotics and Automation (ICRA)*, 2011, pp. 2065-2070.

[4] T. Moranduzzo and F. Melgani, "Automatic car counting method for unmanned aerial vehicle images," *IEEE Transactions on Geoscience and Remote Sensing,* vol. 52, pp. 1635-1647, 2014.

[5] F. Shi, Y. Xi, X. Li, and Y. Duan, "An automation system of rooftop detection and 3D building modeling from aerial images," *Journal of Intelligent & Robotic Systems,* vol. 62, pp. 383-396, 2011.

[6] E. Pakizeh and M. Palhang, "Building detection from aerial images using Hough transform and intensity information," in *Electrical Engineering (ICEE), 18th Iranian Conference on*, 2010, pp. 532-537.

[7] D. Lowe, "Distinctive Image Features from Scale-Invariant Keypoints," *International Journal of Computer Vision,* vol. 60, pp. 91-110, 2004.

Discriminação da densidade vegetacional em ambiente militar

Miguel Gonçalves[1]
08218.goncalves@academiamilitar.pt

José Bioucas-Dias[2]
bioucas@lx.it.pt

José S. Silva[1]
jose.silva@academiamilitar.pt

Jorge Leitão[1]
08152.leitao@academiamilitar.pt

[1] Departamento de Ciências e Tecnologias de Engenharia
Academia Militar
Lisboa, Portugal

[2]Departamento de Engenharia Eletrotécnica e de Computadores
Instituto Superior Técnico
Lisboa, Portugal

Resumo

Este trabalho apresenta um estudo sobre a classificação da densidade vegetacional. Foi utilizado um método de seleção de características, designado *Back Feature Selection (BFS)*, e foram implementados quatro classificadores: *Sparse Multinomial Logist Regression* (SMLR), *K-Nearest Neighbor* (KNN), *Linear Classifier using Principal Component expansion on the joint data* (PCLDC) e *Multi-Class Support Vector Machine* (MSVM). O classificador que produziu o melhor resultado foi o SMLR obtendo um desempenho de 85% para um conjunto de 20 características selecionadas automaticamente.

1 Introdução

Nos processos de decisão militar, o terreno é um fator determinante na forma como decorrem as operações militares. O conhecimento do terreno pode ser decisivo para o sucesso de uma missão permitindo obter vantagens sobre o inimigo. Selecionar a camuflagem correta, definir corredores de mobilidade para o ataque ao inimigo, levá-lo até terrenos onde percam mobilidade são estratégias que se podem usar quando se detém o conhecimento do terreno. Neste sentido, o conhecimento da densidade vegetacional numa determinada zona é uma mais-valia para o processo de tomada de decisão. Face a esta necessidade, é apresentada neste trabalho uma solução para a classificação da densidade vegetacional em ambiente militar. Assim distinguiram-se três classes: vegetação densa, vegetação esparsa e vegetação escassa.

Na classificação da vegetação, a primeira tarefa consiste na extração e seleção de características das imagens seguindo-se a sua classificação. Li [1] sugeriu um método de fusão de características, cor e textura, com o objetivo de classificar três espécies de vegetação diferentes, utilizando imagens aéreas de alta resolução espacial. Esta metodologia com recurso a um classificador MSVM obteve 83,5% de *Overall Accuracy* (OA).

Chapman [2] propôs uma solução para caracterização da vegetação através de fotografias aéreas com informação na banda do visível e imagens com informação na banda do infravermelho. Aplicou um método baseado em árvores de regressão e classificação (CARTs) e o classificador *Random Forest* com o qual obteve 96,8% de classificações corretas.

Wang [3] apresentou um trabalho para delimitar as fronteiras das florestas a partir de imagens aéreas. O método utiliza o algoritmo JSEG (algoritmo de segmentação) que recolhe características *Green Vegetation Index* (GIV) e *Gabor wavelet texture*. Com recurso a um classificador KNN obtém 95 % de OA, tendo a particularidade de a aprendizagem não ser supervisionada.

Neste trabalho pretende-se analisar a influência da seleção e do número de características no resultado da classificação, bem como estabelecer uma comparação entre a performance de vários classificadores.

2 Metodologia

2.1 Dados utilizados

As 5 imagens aéreas ortorrectificadas cedidas pelo Instituto Geográfico do Exército (IGOE) representam uma vasta área da zona militar de Mafra e contêm informação nas três bandas RGB (*red, green e blue*), com um tamanho médio de 12000 por 12000 pixéis, sendo que cada pixel representa 8 centímetros no terreno. Estas imagens foram posteriormente reduzidas em 90% por razões computacionais.

Foram definidas classes (vegetação densa, vegetação esparsa e vegetação escassa) e identificadas as regiões de interesse, que caracterizassem as classes definidas. As regiões de interesse são imagens com tamanho de 100 pixéis por 100 pixéis (superROI's) (retiradas das imagens fornecidas após redução das dimensões) e são utilizadas para a constituição do conjunto de treino usado para a classificação de todas as imagens de teste.

2.2 Métodos

2.2.1 Extração, seleção e normalização de características.

A extração de características é um dos processos mais importantes nos problemas de classificação. Foram extraídas 278 características nas imagens que constituem o conjunto de treino assim como nas imagens que representam o conjunto de teste. Para que as características retiradas expressassem corretamente a densidade de vegetação verificou-se, após análise, que calcular as características das superROI's em regiões (ROI´s) de 50 por 50 pixéis seria a abordagem mais adequada. Estas foram parametrizadas com 10 características de primeira ordem e 3 de segunda ordem.

Foi implementado um método de seleção de características disponibilizadas na *toolbox Partern Recognition Tools (PRTools)* [4], denominado *Back Feature Selection* [5].

Verificou-se que o intervalo dos dados que constituem o conjunto de treino é extenso, como tal, realizou-se a normalização de todas as características.

2.2.2 Classificação

Utilizaram-se métodos de classificação multi-classe supervisionada para se realizar a classificação do coberto vegetal. O objetivo é parametrizar vários classificadores que, dado um conjunto de treino com *n* amostras das três classes, seja capaz de distinguir as diferentes classes perante uma nova imagem. Implementaram-se os classificadores: *Sparse Multinomial Logist Regression* [6], *K-Nearest Neighbor*, *Linear Classifier using Principal Component expansion on the joint data* [4] e um classificador *Multi-Class Support Vector Machine* com utilização de duas funções de *kernel* diferentes: *Radis Basis Function* (RBF) e *Multilayer Perceptron* (MPL).

Uma característica comum à maioria dos classificadores e que influência diretamente o seu desempenho é a existência de um ou mais parâmetros que condicionam a forma como o classificador aprende e classifica as amostras de teste. Para estimar os parâmetros de entrada dos classificadores utilizados utilizou-se uma técnica designada *10-fold cross validation*. Esta técnica permite encontrar os parâmetros ótimos para cada classificador com base na minimização dos erros de classificação do conjunto de treino.

3 Resultados

Com o objetivo de analisar a influência do número de características no resultado da classificação, bem como estabelecer uma comparação entre a performance de vários classificadores, utilizaram-se 5 imagens de teste diferentes (perfazendo um total de 959 ROI's analisadas) e registaram-se para cada classificador os valores de OA médios.

O número de características é um fator determinante para o sucesso de um problema de classificação. Neste trabalho foi utilizado um número de características entre 5 e 40 com incrementos de 5. Verificou-

se que considerar mais de 40 características ocorre a uma diminuição da OA em todos os classificadores, sendo esse o motivo pelo qual a análise foi realizada até as 40 características.

Com o classificador SMLR a utilização de 20 características foi suficiente para obter 85% de OA. O parâmetro de regularização λ determinado recorrendo a técnica *10-fold cross-validation* utilizado foi λ=0,78.

Classificaram-se as mesmas imagens com o classificador KNN, para encontrar os parâmetros que conduziam a melhor classificação, tendo sido testado como métricas de distância as seguintes variantes: distância euclidiana, distância de *Chebycheb*, correlação e distância de *Mahalanobis*. Com este classificador, o melhor valor de OA obtido foi de 73,2%, calculado pela métrica de correlação, considerando um valor de k=3 e recorrendo às 15 características mais importantes na distinção das classes.

O *Linear Classifier using Principal Component expansion on the joint data (PCLDC)* da *toolbox PRTools* [4] foi utilizado na classificação das imagens de teste. Este classificador permite definir qual o número de vetores próprios (N) utilizados para projeção dos dados do conjunto de treino. Com N=2, obtido pela técnica *10-fold cross validation*, foi possível obter um valor de OA de 77,1% quando se utilizavam 5 características.

Por último, com o classificador MSVM realizaram-se testes no sentido de encontrar qual a função de *kernel* que permitia obter melhores resultados. O melhor desempenho deste classificador verificou-se com a utilização de 25 características com uma função de *kernel* RBF (parâmetros σ =4 e C=16,6) com um OA de 76%.

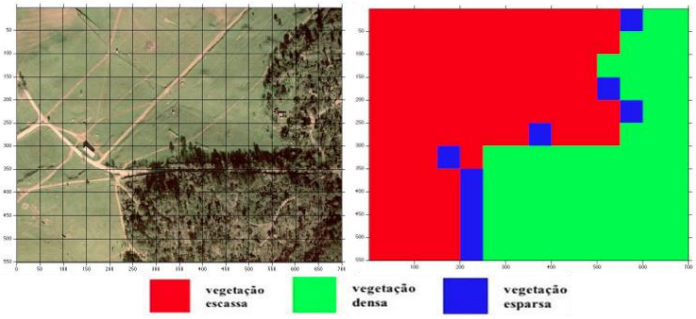


Figura 1-Imagem real (à esquerda) e imagem artificial (à direita) produzida com o resultado do classificador SMLR.

A Figura 1 mostra à esquerda a imagem original de teste, e à direita uma imagem artificial construída com base no resultado do classificador SMLR para uma das imagens de teste (550 por 700 pixéis). Neste exemplo conseguiu-se uma percentagem de classificações corretas de 94,8%.

Realizou-se uma comparação entre os melhores resultados (valores médios) obtidos pelos classificadores em função do número de características, sendo o resultado apresentado no gráfico seguinte.

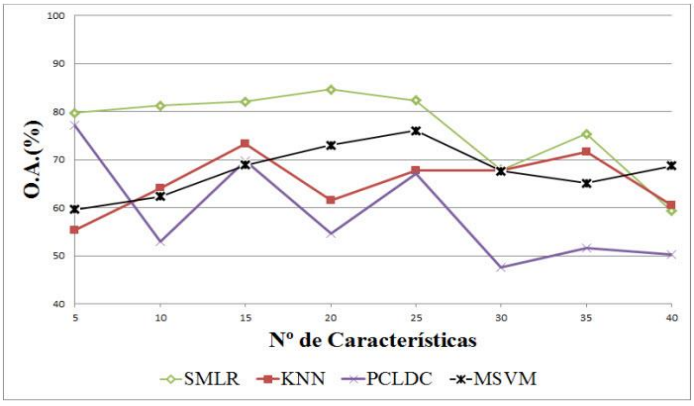


Gráfico 1- Comparação dos resultados obtidos com os vários classificadores em função do número de características.

Importante referir que no gráfico anterior, os valores referidos para o classificador KNN foram conseguidos com a utilização da métrica correlação, visto ser a que permite obter melhores resultados na maioria das situações. Já para o classificar MSVM foi utilizado a função de *kernel* RBF pelo mesmo motivo.

Por análise gráfica verificou-se que o classificador SMLR é o que demostra uma variação menor ao aumento do número de características, mantendo uma constante superioridade relativamente aos restantes. Para um número de características entre 15 e 25 ocorre o melhor desempenho do classificador, obtendo um valor de 84,6% (para 20 características e com λ =0,78).

Realizada a comparação entre classificadores, resta ainda verificar qual a influência da utilização de um seletor de características nos resultados. A utilização de um seletor de características (BFS) conduziu a um melhor desempenho de todos os classificadores. A tabela seguinte mostra os melhores resultados dos classificadores com e sem a utilização do seletor de características.

	Sem seletor de características (278 características)	Com seletor de características		Aumento da performance (%)
	OA(%)	OA(%)	Nº Características	
SMLR	60,5	85,0	20	24,5
KNN	62,7	73,2	15	10,5
PCLDC	51,3	77,2	5	25,9
MSVM RBF	56,7	76,0	25	19,3

Tabela 1- Comparação entre os resultados dos classificadores com e sem a utilização de um seletor de características.

4 Conclusões

A extração de um elevado número de características permite que sejam usados algoritmos de seleção para eliminar características redundantes e diminuir a complexidade do modelo de representação do conjunto de treino, uma vez que reduz a dimensionalidade do espaço de características, correspondendo a um aumento significativo da performance, na ordem dos 20% em todos os classificadores exceto no KNN onde se regista um aumento de 10%.

Este trabalho também permitiu concluir que é no intervalo entre as 15 e as 35 características que a maioria dos classificadores obtém as melhores classificações, com exceção apenas para o PCLDC que permite obter melhores classificações com um número de características mais reduzido (5 características). Verificou-se que o melhor classificador é SMLR, sendo o único que permite obter 85% das classificações corretas. Neste caso as características não são linearmente separáveis, advindo daí a necessidade de utilizar classificadores com funções de *kernel*. As funções de *kernel* permitem a projeção das amostras para um espaço de características diferentes do original, onde as amostras das várias classes podem ser divididas linearmente. É esta versatilidade que está na base dos resultados produzidos pelos classificadores SMLR, e SVM RBF.

Agradecimento

Ao CINAMIL, Centro de Investigação da Academia Militar, através do projeto FUSIMIL, pelo financiamento que suportou as despesas relativas à conferência RECPAD'2014.

Bibliografia

[1] R. Hayward, R. Walker, Z. L. Zhengrong Li, Y. L. Yuee Liu, et al., "Color and texture feature fusion using kernel PCA with application to object-based vegetation species classification," 2010, pp. 2701-2704.

[2] D. S. Chapman, A. Bonn, W. E. Kunin, and S. J. Cornell, "Random Forest characterization of upland vegetation and management burning from aerial imagery," Journal of Biogeography, vol. 37, pp. 37-46, 2010.

[3] Z. Wang, R. Boesch, and C. Ginzler, "Forest delineation of aerial images with Gabor wavelets," International Journal of Remote Sensing, vol. 33, pp. 2196-2213, 2012.

[4] R.P.W. Duin, P. Juszczak, P. Paclik, E. Pekalska, D. de Ridder, D.M.J. Tax, S. Verzako, PRTools4.1, A Matlab Toolbox for Pattern Recognition, Delft Universityof Technology, 2007.

[5] I. Guyon and A. Elisseeff, "An introduction to variable and feature selection," J. Mach. Learn. Res.,vol. 3, pp. 1157–1182, Mar. 2003. [Online]. Available: http://dl.acm.org/citation.cfm?id=944919.944968

[6] J. Bioucas-Dias and M. Figueiredo, "Logistic regression via variable splitting and augmented Lagrangian tools," Instituto Superior Técnico, TULisbon, Lisbon, Portugal, 2009, Tech. Rep.

Metadata Extraction from Ancient Documents

Henrique Mata[1], João R. Caldas Pinto[1] and Maria C.S. Nunes[2]
henrique.mata@tecnico.ulisboa.pt, joao.c.pinto@tecnico.ulisboa.pt, mcarmonunes15@gmail.com

[1]IDMEC/ LAETA, Instituto Superior Técnico, Universidade de Lisboa

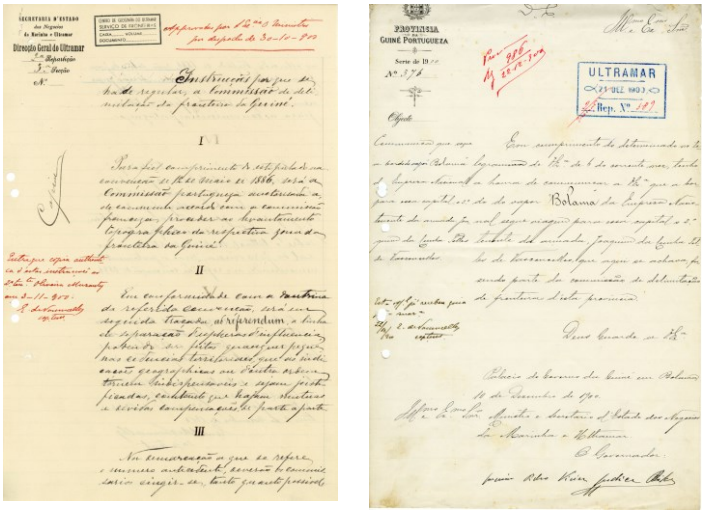
[2] IICT, Instituto de Investigação Científica Tropical, Travessa Conde da Ribeira, 9, B, 1º, Lisboa, Portugal

Abstract

The problem addressed in this work is the extraction of descriptive metadata from handwritten and typed documents, comprising the recognition of artifacts such as stamps, underlines and annotations, and also the recognition of tables, through the use of thresholding and morphologic operations with very promising results.

1 Introduction

Historical document collections residing in libraries, museums and archives are currently being digitalized for preservation purposes and to make them available worldwide through large on-line digital libraries. However, efforts should also focus on technologies aimed at reducing the human effort required for the annotation of the raw images with informative content. The presence of artifacts, such as stamps, underlines, annotations may indicate the presence of relevant information. In this sense, the extraction of metadata from this type of information constitutes an ultimate goal for institutions like libraries, motivating the building of an automatic and efficient recognition system [6].



a) b)

Fig. 1 – a), b) Examples of ancient documents

The presented ideas will be applied to a representative set of documents from the IICT Library on themes concerning the borders of Guinea Bissau. These documents are manuscripts and typed scripts from the late nineteenth and early twentieth century. Two images of this set are shown in figure 1.

2 Related Techniques

Different work has been done in regard to the area of document image analysis [1],[2],[3]. Muge et al. [1] proposed a methodology for automatic extraction and recognition of features from books of the Renaissance. Using grey level images, the approach consists in a page segmentation followed by a mathematical morphology methodology. The concept guiding the segmentation methodology refers to a topographic analogy. The signs (text and figures) to detect are like "valleys" or "furrows" (darker regions) carved on a basically plane background (lighter regions). Based on the segmented pages, the separation of non-text (figures, labels and illuminated letters) from text characters is achieved through the use of mathematical morphology. Fundamentally, the algorithm uses the main directions in which the image's elements evolve, i.e., assuming that the text has an orientation of 0º (length) and 90º (height) in the majority of lines composing the characters.

3 Implemented Techniques

In the proposed algorithm, the recognition of artifacts takes advantage of the fact that, most of the times, artifacts present a different colour from the text, which motivates the use of colour thresholding techniques to separate non-text contents, such as tables, stamps, annotations, underlines, from text. In order to simplify the colour thresholding expression, the conversion from RGB to HSV is performed [5]. The reason behind the selection of the HSV color channel is related with the hue component, that gives the dominant wavelength of a colour, allowing to identify more easily a certain color. Using the binary images obtained with the thresholding techniques, a mathematical morphology based algorithm follows next. Fundamentally, this algorithm seeks for the straight lines that characterize each artefact. It assumes that, in a document, tables and stamps are the only artifacts showing vertical lines, while underlines, with the exception of tables and stamps, are the only artifacts lines exhibiting horizontal lines. To retrieve the morphological features sought, an erosion process takes place. In order to guarantee that only areas belonging to the artifacts are considered, an area opening operation is executed to remove residual areas. The final criterion to infer about the presence of an artifact is an area threshold that considers that an artifact is present if the number of pixels in the resulting image is equal or higher than the fixed threshold. For stamps, in addition to this criterion, the number of vertical lines has to equal to two. Regarding tables, the number of both vertical and horizontal lines has to be equal or higher than two.

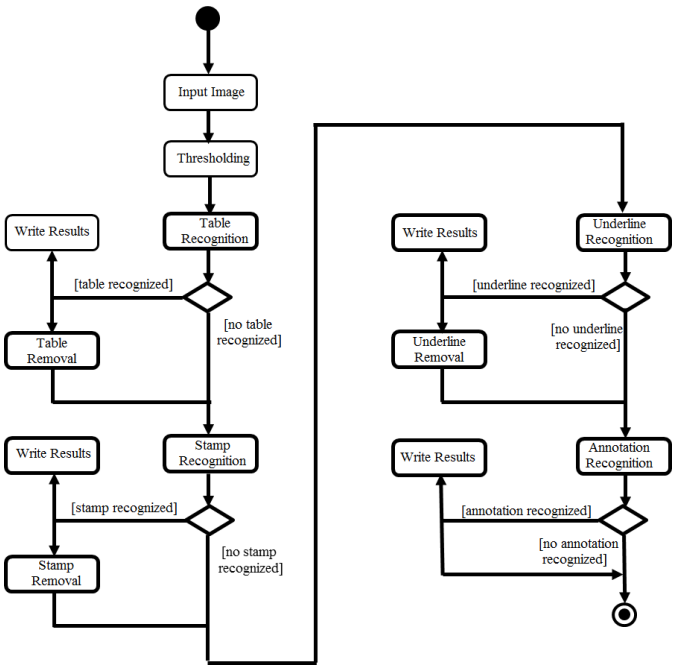
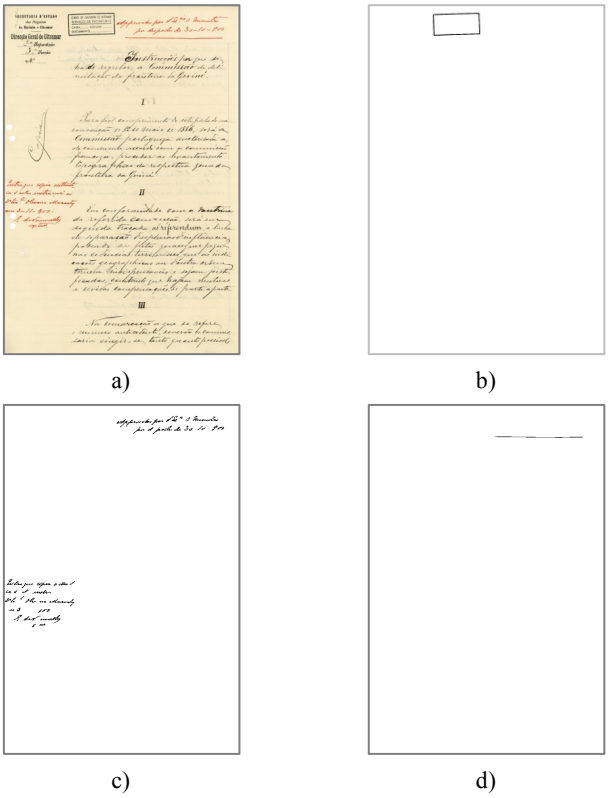


Figure 2 - Activity diagram of the artifact recognition algorithm

However, due to the common morphological key features that some artifacts share, artifacts may be incorrectly classified. To avoid this, after the recognition of an artifact, the result obtained in the residual area removal stage is used in a morphologic reconstruction operation to find the rest of the artifact, in order to remove it from the binary image that is going to be used to recognize the next artifact, eliminating possible interferences. The sequence of artifacts recognized is the following:

tables, for having longer lines than stamps, are the first artifacts to be identified. After being recognized, these are removed from the image, leaving stamps as the only artifacts exhibiting vertical lines. Stamps follow next in the recognition process. Once again, after being identified, these are also removed, due the presence of horizontal underlines present in the top and bottom of the stamps that could be understood as underlines in the underline recognition process. Finally, after the recognition of the tables, stamps and underlines, the only artifacts remaining should be the annotations. Due to the arbitrary shape that these can take, it is assumed as annotations any coloured region that occupies an area higher number than a fixed threshold. Thus, the recognition of annotations consists in the removal of all the artifacts recognized, followed by an area threshold. The activity diagram of this algorithm is illustrated in figure 2. The application of this algorithm is presented in figure 3, where it is possible to see the recognition of the different artifacts.



a) b) c) d)

Fig. 3 – a) Original Image b) Stamp recognized c) Annotations recognized d) Underline recognized

Results Discussion

In order to test the proposed techniques, the new approach was applied to the set of 122 documents. The results may be observed in Table 1 that results from the confusion matrix [4]. The results of table recognition were conditioned by the low number of tables present in the set analysed. Moreover, part of these had their borders defined with dashed lines, which drive the algorithm to fail. The lines that define the tables are usually black, which makes their recognition to rely exclusively on the morphologic stage. The mask used in the erosion process spared lines from margins and from annotations with vertical strokes, which were after considered as being part of tables, leading to inaccurate classifications. To avoid this, a criterion regarding the number of vertical and horizontal lines was implemented in decision stage to classify an image as containing tables. To what concerns the stamps, the presence of annotations with vertical straight lines was found to be one of the main causes of inaccuracy in the recognition of this type of artifacts. However, most of the inaccurate classifications were due to the fading of stamps. A foreseen drawback in the morphological stage was related with the rotation that stamps might present. However, the mask used was able to tolerate considerable rotations. To what concerns the recognition of underlines, this was mostly conditioned by their consistency. The pressure variations that occur in the underlining process conduct to a fading of some parts of the underline, that is critical when occurring in the interior of the underlines, due to possible disintegration of the underline in the thresholding stage and further removal of the entire underline in the residual area removal stage. In the documents analysed, the annotations usually appear in red. Due to the red appearing in the extremes regions of the hue channel, its identification is easily processed, allowing an accurate thresholding expression and therefore an accurate recognition of all the red coloured artifacts. However, not all red areas are related to these artifacts. Some documents present red typed characters in headings, headlines, titles and others that are part of the original document. In this sense, although they were not introduced after the creation of the document, as a regular annotation, they still point relevant content, which led to consider those as annotations.

4 Conclusions and future work

The algorithm presents a high degree of generalisation as the current experiments indicate, even though the differences present in the set of images analysed. It was relatively easy to detect the existence of underlines, stamps and annotations using an adequate analysis of the HSV components of the image, together with basic morphological operations. However, the task of doing artifact recognition without the aid of the 'semantic level' is problematic. As future work, the recognition of the several artifacts may include a stage where the document is segmented in text and non-text areas, restricting the search of underlines and annotations to where they are expected to appear. It would be also useful to automatically distinguish manuscripts and typed scripts.

Table 1 – Results of the Artifact Recognition Algorithm

%	Tables	Stamps	Underlines	Annotations
Precision:	87,00	84,61	88,89	91,53
Sensibility:	77,50	89,19	93,02	93.10
Specificity:	99,12	92,94	72,22	92,19
Accuracy**:**	97.54	91.8	86.89	92.62

Acknowledgments

This work is funded by National Founds through FCT - Foundation for Science and Technology (Fundação para a Ciência e Tecnologia) under the Project PTDC/ATP-GEO/4645/2012,IDMEC.

References

1. F. Muge, I. Granado, M. Mengucci, P. Pina, V. Ramos, N. Sirakov, J.R. Caldas Pinto (2000). "Automatic Feature Extraction and Recognition for Digital Access of Books of the Renaissance", ECDL 2000, pp. 1–13, Springer-Verlag Berlin Heidelberg 2000.
2. João R. Caldas Pinto, Lourenço Bandeira, João M. C. Sousa, Pedro Pina, (2005), "Combining Fuzzy Clustering and Morphological Methods for Old Documents Recovery", 2nd Iberian Conference on Pattern Recognition and Image Analysis (IbPRIA05), June 7-9, 2005, Estoril, Portugal. Lectures Notes in Computer Science, LNCS 3523, pp. 387–394, Springer-Verlag, Berlin Heidelberg 2005.
3. Pedro Castro, Caldas Pinto, J. R. (2007). "Methods for Written Ancient Music Restoration", ICIAR 2007, Canada 18-20, Montreal, Canada, August 2007, LNCS 4633, pp 1194–1205, Springer-Verlag Berlin Heidelberg, 2007.
4. Michael R. Berthold, Christian Borgelt, Frank Hoppner, Frank Klawonn (2010), Guide to Intelligent Data Analysis: How to Intelligently Make Sense of Real Data, Springer.
5. Gonzalez, R. C., Woods, R. E., & Eddins, S. L. (2009). Digital Image Processing Using Matlab. Gatesmark Publishing.
6. Fichter, Darlene (1999). Administrative and Factual Metadata for Intranets-Issues and Options, Online , Vol. 23, No. 6, 1999.

Attenuation of speckle noise in OCT images

Fabiana Rodrigues
José Silvestre Silva
Jaime Santos

Centro Cirúrgico de Coimbra
Departamento de Ciências e Tecnologias de Engenharia – Academia Militar
Departamento de Engenharia Eletrónica e Computadores da Universidade de Coimbra

Abstract
The speckle noise causes a significant degradation of image quality, interfering negatively in its processing. In this work several filters were tested and statically evaluated, computing their performance. The results of this study show that SureShrink and wiener filters have the best performance.

Introduction
The speckle noise is a multiplicative noise that results from the interference of coherent waves reflected in anatomical structures. This type of noise obeys to the Rayleigh distribution and it can be described by the following expression[1]:

$$I(x,y) = f(x,y)n(x,y) \quad (1)$$

where $I(x,y)$ is the image; $f(x,y)$represents the image without noise and $n(x,y)$ is a speckle noise.

There are several types of filters used to reduce the speckle noise in Optical Coherence Tomography (OCT) images without blur the boundaries and without losing much details in the images, such as, adaptive filters, anisotropic diffusion filters; wavelet based technique and nonlinear anisotropic filters.

The *Lee* [2], *Frost* [1] *wiener* [3], *Kuan* [4] filter are some examples of the adaptive filters. They are based on local statistics (local variance and local mean) around the central pixel. Some of these methods, like wiener filter, are effective in reducing speckle noise, improving contrast and image quality. However the Lee and Frost filter tend to over smooth the boundaries, losing significant details contained in the images which is a disadvantage for clinical analysis and for image processing [5,6].

Other technique based on local statistics is the linear-least squares estimation method. Nevertheless, this method provides an insufficient noise reduction [6]. To overcome this limitation Wong [6] suggested a noise attenuation method using a Bayesian linear least squares estimators. This algorithm reveals the noise distribution of the observed data and provides a good speckle reduction. It was tested in mouse retinas, showing a clear preservation of its small features.

There are two types of filters that have been heavily studied for speckle noise reduction. They are anisotropic diffusion-based methods and wavelet-based methods. The first study about anisotropic diffusion filtering was developed by Perona and Malik [7]. However this denoising technique has some disadvantages, such as the staircasing effect, which results in shape fluctuations of the edges, because in smooth step boundaries is created regions in image separated by jumps [8 In the wavelet-based strategies the images are decomposed in its frequency components and in its coefficients being associated with a speckle noise and they are eliminated by a threshold [7]. There are several types of threshold techniques but the most used are VisuShrink [9], SureShrink [10] and BayesShrink [11]. Several studies reported that VisuShrink over smooth the image because it is based on a Universal Thresold and that doesn't fit to discontinuities in the image [9]. However, the BayesShrink has the best performance.

In this work several speckle-reduction filters were implemented to smooth the speckle noise in OCT images, comparing their performance.

Speckle – reduction filters
Winer filter: It is based on local statistic (mean and variance) with aim to minimize the mean square error (MSE) between desired image and smoothed image. This method preserves the boundaries and others regions with high frequency [3].
Mediana filter: It is a nonlinear filtering technique where the central pixel of the window is replaced by a median value of neighbouring pixels [12].
Lee filter: This method assumes that noise is multiplicative with a Gaussian distribution. It is based on local and linear minimization of the mean square error (LLMSE) [2].
Frost filter: The filtered pixel is replaced by a value calculated based on distance from the pixel of interest and local variance [1].
Perona-Malik Anisotropic Diffusion (PMAD): It was the first work in this area and it characterized by introduction of diffusion coefficient [7].
Bilateral filter: The bilateral filter combines two types of information: photometric similarity and geometric closeness. The central pixel is replaced by a weighted average of neighbouring pixels where the weights depend on neighbourhood intensity and neighbourhood color. The intensity peaks will have lower weight, even when they are close to central pixel [13].
Wavelet-based methods: The image is decomposed in subbands, namely the HH, HL, LH (using high-pass filter) and LL (using low-pass filter) bands. The LL band corresponds to the approximation coefficients and the others three correspond to the detail coefficients. This decomposition enables us to study the signal in different frequency bands. Then an appropriate threshold is selected to remove small coefficients that occur most likely from noise signals. There are several thresholding wavelet techniques but the most used are VisuShrink [9], BayesShrink [11] and SureShrink [10].

Speckle – reduction performance metrics

Signal-to-noise ratio (SNR): SNR evaluates the ability to suppress speckle noise. A more efficient proposed method presents the greater SNR value. It is expressed in decibels and it is defined by [14]:

$$SNR = 10\ \log_{10} \frac{\sum_{x=0}^{M-1}\sum_{y=0}^{N-1}[f(x,y)]^2}{\sum_{x=0}^{M-1}\sum_{y=0}^{N-1}[f(x,y)-g(x,y)]^2} \quad (2)$$

where $f(x,y)$ is the original image; $g(x,y)$ is the filtered image and MN represents the image size.
Root mean Square Error (RMSE): It indicates the similarity between the original image and filtered image. The lower value indicates that there is a higher similarity between images [6].

$$RMSE = \sqrt{\frac{1}{MN}\sum_{x=0}^{M-1}\sum_{y=0}^{N-1}[f(x,y)-g(x,y)]^2} \quad (3)$$

Coefficient of Correlation (CoC): The CoC is an indicator of the similarity of filtered image and original images. Its value is 1 in ideal case and 0 for uncorrelated images. The larger CoC value, the lower the degradation of the edges caused by filter [6].

$$CoC = \frac{\sum((g(x,y)-\overline{g(x,y)}) \times (f(x,y)-\overline{f(x,y)}))}{\sqrt{\sum(f(x,y)-\overline{f(x,y)})^2 \times \sum(g(x,y)-\overline{g(x,y)})^2}} \quad (4)$$

where $\overline{f(x,y)}$ and $\overline{g(x,y)}$ represent the mean of respective images.

Data Set
To compare these filters, we used artificial images contaminated with different noise levels, simulating the real images. In this study, the variances used were 0.01, 0.03 and 0.05. I tried to use images simulated the complexity and variability of OCT images, presenting: (i) similar regions not being possible to see clearly the boundaries between structures; (ii) regions with high contrast and (iii) small structures, analysing the changes in their shape, size and position. In a second step

of study were used 414 OCT images, in TIF format and with 1024 x 496 of resolution.

Figure 1. Example of an artificial image used in this study.

Results & Discussion

It is only shown the results for image contaminated with variance 0.05, however they are reproducible for other variances.

	σ^2 =0.05		
Filter	**SNR**	**CoC**	**RMSE**
Image without noise	-	**1.00**	**0.00**
Image with noise	**13.26**	**0.89**	**28.14**
Wiener	18.08	0.96	16.16
Mediana	16.11	0.93	20.27
Lee	17.75	0.95	16.78
Bilateral	15.47	0.92	21.80
Frost	17.50	0.95	17.27
PMAD	17.95	0.95	16.38
***SureShrink* (2 levels)**	18.88	0.96	14.73
***SureShrink* (3 levels)**	19.23	0.97	14.15
***BayesShrink* (2 levels)**	17.37	0.95	17.53
***BayesShrink*(3 levels)**	17.53	0.95	17.20
***VisuShrink* (2 levels)**	17.29	0.95	17.69
***VisuShrink* (3 levels)**	16.78	0.93	18.75

Table 1. Comparison of various filters (for image contaminated by speckle noise of variance 0.05).

The results show that:

- The Lee, Frost and Wavelet based on Universal threshold (VisuShrink) filters, reduce the speckle noise. However, the intensity transitions aren't preserved, losing details and having degradation of the edges. These results are confirmed by low value of CoC;
- The median and bilateral filters present a low noise suppression in homogeneous regions, reducing image sharpness and hindering the perception the texture of region. These limitations are confirmed by low values of SNR and by high values of RMSE. The reduced efficiency of the bilateral filter is due to its mathematical formalism that assigns a lower weight to the intensity peaks;
- The wiener filter and Sureshrink filter have the best performance, as seen in the statistical metrics (SNR, RMSE, CoC). They remove the noise without adding artifacts and preserve the small details of the image;
- The introduction of the diffusion coefficient defined by Perona and Malik, allowed a selective smoothing, with greater diffusion in homogeneous regions, and a lower diffusion in the contours of the image;
- The wavelet filter based on BayesShrink threshold demonstrates a reduced ability to remove the noise (low values of SNR, RMSE and CoC). However its efficiency isn't optimized by increasing the number of decomposition levels

Similar results were obtained when the filters were tested in OCT images. The wiener filter and Sureshrink filter have the best performance, increasing the quality of images.

References

[1] A. Cameron, D. Lui, A. Boroomand, J. Glaister, A. Wong e K. Bizheva, "Stochastic speckle noise compensation in optical coherence tomography using non-stationary spline-based speckle noise modelling," Biomedical Optics Express, vol. 4, n.º 9, pp. 1769-1785, 28 Agosto 2013.

[2] A. Wong, A. Mishra, K. Bizheva e D. A. Clausi, "General Bayesian Estimation for speckle noise reduction in optical coherence tomography retinal imagery," Optics Express, vol. 18, n.º 8, pp. 8338-8352, 6 April 2010.

[3] P. Serrannho, C. S. T. Maduro, J. Cunha-Vaz e B. R., "Synthetic OCT data for image processing performance testing," 18th IEEE International Conference on Image Processing (ICIP), pp. 401-404, 11-14 September 2011.

[4] Soni e L. Bhaiya, "A novel threshold selection technique for speckle filtration from medical images," lnternational Journal of Scientific Engineering and Technology, vol. 1, n.º 4, pp. 42-45, October 2012.

[5] K. Prakash, B. e B. VenuGopal, "Image independent filter for removal of Speckle noise," International Journal of Computer Science Issues, vol. 8, n.º 5, pp. 196-201, September 2011.

[6] K . Sivakumar, M. Gayathri e D. Nedumaran, "Speckle Filtering of Ultrasound B-Scans Images - A Comparative Study of Single Scale Spatial Adaptive Filters, Multiscale Filter and Diffusion Filters," International Journal of Engineering and Technology, vol. 2, n.º 6, pp. 514-523, 6 December 2010.

[7] K. Kuan, A. Sawchuk, T. Strand e P. Chavel, "Adaptive Noise Smoothing Filter for Images with Signal-Dependent Noise," *IEEE Transactions on Pattern Analysis and Machine Inteligence*, Vols. %1 de %2PAMI-7, n.º 2, pp. 165-177, March 1985.

[8] K. Karthikeyan e C. Chandrasekar, "Speckle Noise Reduction of Medical Ultrasound Images using Bayesshrink wavelet Threshold," *International Journal of Computer Applications*, vol. 22, n.º 9, pp. 8-14, May 2011.

[9] A. Budes, B. Coll e J. Morel, "The staircasing effect in neighborhood filters and its solution," *IEEE Transactions on Image Processing*, vol. 15, n.º 6, pp. 1499-1505, June 2006.

[10] J. Weickert, "A review of nonlinear diffusion filtering," *Scale-Space Theory in Computer Vision Lecture Notes in Computer Science* , vol. 1252, pp. 1-28, 1997.

[11] M. Nitzberg e T. Shiota, "Nonlinear image filtering with edge and corner enhancement," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 14, n.º 8, pp. 826-833, August 1992.

[12] T. Joel e R. Sivakumar, "Despeckling of Ultrasound Medical Images: A Survey," *Journal of Image and Graphics*, vol. 1, n.º 3, pp. 161-165, September 2013.

[13] C. Prabhakar e K. Praveen, "An Image Based Technique for Enhancement of Underwater Images," *International Journal of Machine Intelligence*, vol. 3, n.º 4, pp. 217-224, 2011.

[14] R. Vanithamani e G. Umamaheswari, "Performance Analysis of Filters for Speckle Reduction in Medical Ultrasound Images," *International Journal of Computer Applications*, vol. 12, n.º 6, pp. 23-27, December 2010

Safety Level opposite Leakage in underground fuel tank based on the Model of Pattern Recognition

Erico Gaspar Lisboa
http://paginas.fe.up.pt/~shrha/index.html

PhD Student (CNPq/CsF Fellowship, Brazil)
Faculty of Engineering, University of Porto, FEUP

Abstract

The present work consisted in application of the model pattern recognition (MPR) of a pattern safety. This kind of strategy was adapted to protection and management of groundwater resources opposite a possible aquifer contamination adopting a safety level against leakage (SLL) in underground fuel tank.

1 Introduction

The failure probability in underground fuel tank could lead the leakage and contamination the soil and water, in another word an environment threat. This kind of threat could be characterized by dermal contact and ingestion of contaminated water with the main component of fuel: BTEX (benzene, toluene, ethylene and xylene). Whatever the contact way with these components grows the hypothesis of appear the effects adverse: toxicity and carcinogenicity.

Almost all these problems have originals in infrastructure of the fuel station, kind of tank materials (KTM), monitoring of a possible leakage (MPL) and time operating (TOP), namely the susceptibilities criteria. These criteria can suggest a pattern classification of security opposite a leakage. When this kind of classification is associated to vulnerability conceptions can be translated in risk conception. On other hand there are a few works in expert literature that proposed this kind of association. This context the proposal to measure a ranking between the vulnerability of underground environment and susceptibility criteria was mooted by [1]. However, this kind of proposal not gave emphasis to relative importance each-one criteria and there'd a way to classify a pattern SLL of each the fuel station.

The aim this article was build a pattern safety matrix (PSM) considering the susceptibilities criteria and natural vulnerability of underground environment. These criteria were categorized in according with hazard degree after insert a fuel tank in underground environment. The pH was a criterion utilized to [1] and held-on like category. On other hand, the vulnerability was analysed by GOD index model according the proposal by [2]. This index has been dependant of an interactive fuzzy combination process between three hydrogeology variables. Thus, was introduced a fuzzy inference system (FIS) with support of geostatic analysis, namely the ordinary kriging. The FIS was processed by fuzzy inference model Mamdani.

The fuzzy natural vulnerability index (NVI*f*) as final product of FIS was obtained by inverse operation of logic fuzzy by gravity center method based on the proposal of [3]. The interpolation of NVIf degree results was adjusted to spherical (SHP), exponential (EXP) and gaussian (GAU) models. The model with better performance was chosen for two processes: sum of squared residuals (SSR) and multiple correlations (R^2). In sequence this model was validated through by cross validation techniques. In addition based on the original proposal in [1] was considerate that PSM had one more criteria: vulnerability in around of station fuel (VSF).

2 Prescriptions to construction of PSM and SLL

The pattern safety matrix (PSM) was developed with based on susceptibility and vulnerability criteria. The matrix number of columns was associated by the safety level opposite leakage (SLL) that expressed a way to relate the criteria. The levels were patterned in five degree:

Acceptable – **A** ($5 \leq SLL > 4$): The susceptibilities criteria is up 4xSLL less than natural vulnerability of underground environment and in around of fuel station; Moderate – **M** ($4 \leq SLL > 3$): The susceptibilities criteria is up 3xSLL less than natural vulnerability of underground environment and in around of fuel station; Worry – **W** ($3 \leq SLL > 2$): The susceptibilities criteria is up 2xSLL less than natural vulnerability of underground environment and in around of fuel station; Sub-critical – **S** ($2 \leq SLL > 1$): The susceptibilities criteria is up 2xSLL less and equal the natural vulnerability of underground environment and in around of fuel station; Critical – **C** ($SLL = 1$): The susceptibilities criteria is equal the natural vulnerability of underground environment and in around of fuel station. The attribution of SLL number was done by a convention between value 1 - 5 (Table 1).

Table 1 – Pattern safety matrix (PSM).

Criteria	Value and framing pattern of SLL					
	Degree	A	M	W	S	C
	SLL	5	4	3	2	1
Vulnerability	NVI*f*	N	L	M	H	E
	VSF	0	1	1	2	3
Susceptibilities	TOP	1	1	2	2	3
	KTM	1	1	2	3	4
	MPL	1	1	2	2	3
	pH	1	1	2	2	3

This procedure was realized to configure the recognition model to that PSM could be extend to another situation not frame like that pattern. Some this situation could be illustrated: (1) A station fuel located in low vulnerability zone in that susceptibilities criteria adopted aren't sufficient of break a possible leakage; (2) A station fuel located in extreme vulnerability zone in that susceptibilities criteria adopted are sufficient of break a possible leakage.

The NVI*f* degree was obtained through of proposal based on [2] classify as: negligible – N ($0 \leq NVIf \leq 0.10$); low – L ($0.10 \leq NVIf \leq 0.30$); medium – M ($0.30 \leq NVIf \leq 0.50$); high – H ($0.50 \leq NVIf \leq 0.80$) and extreme – E ($0.80 \leq NVIf \leq 1$). On other hand, the VSF value was obtained by class that were prescript between intervals of 0 to 3. The TOP were categorized with base on year of until life: 1 ($TOP \leq 10$ year); 2 ($10 > TOP \leq 20$ year) and 3 ($TOP > 20$ year).

The categories of KTM were framed in four: 1 (steel-carbon without cathodic protection); 2 (steel-carbon with cathodic protection); 3 (double wall without coating) and 4 (double wall with coating). This same way the criteria relate to MPL had three categories: 1 (manual); 2 (manual with another operations) and 3 (automatic). The value of pH was categorized in: 1 ($pH > 6.5$); 2 ($5.5 \leq pH \leq 6.5$) and 3 ($pH < 5.5$).

3 Application a MPR to reading the PSM

The MPR was characterized by transformation of PSM to a way normalized (PSM_N) in according with the prescription of [4]. This process resulted in a PSM_N equal:

$$PSM_N = \begin{bmatrix} 0{,}90 & 0{,}70 & 0{,}20 & 0 \\ 0{,}67 & 0{,}67 & 0{,}33 & 0 \\ 0{,}50 & 0{,}50 & 0 & 0 \\ 1 & 0{,}50 & 0{,}50 & 0 \\ 0{,}67 & 0{,}67 & 0{,}33 & 0 \\ 1 & 0{,}50 & 0{,}50 & 0 \end{bmatrix} = PSM_{N_{i,h}} \quad (1)$$

Where i is the number of criteria and h is associated to safety level degree. The matrix normalized components were obtained through of relation between the difference ($y_{i,h}$ - $y_{i,5}$) and ($y_{i,h}$ - $y_{i,5}$). When this difference is equal 1 the SLL is classified as totally acceptable. If this value is 0 the classification is totally critical. The recognition pattern process was started with the application of equation (2).

$$r_{i,h} = \begin{cases} 0 \\ x_{i,h} - y_{i,5} \\ x_{i,1} - y_{i,5} \\ 1 \end{cases} \quad (2)$$

Where $x_{i,h}$ is relative to sample random data of fuel tank being considered the criteria i ($i = 1, 2, 3, 4, 5, 6$) and safety level degree h ($h = 1, 2, 3, 4, 5$). After this the PSM became a matrix with $r_{i,h}$ value it being possible calculate a weighted distance between the $r_{i,h}$ and PSM_N. The equation (3) related the weight of each i criteria (w_i) that were obtained through of analyse hierarchical based on [5].

$$d_{h,j} = \left\{ \sum_{=}^{6} \left[w_i \cdot \left(x_{i,h} - PSM_{N_{i,h}} \right)^2 \right] \right\}^{\frac{1}{2}} \quad (3)$$

When $d_{h,j} = 0$, the value of $r_{i,h}$ is equal the $PSM_{Ni,h}$. The most important criteria were attributed for NVI*f* following of TOP, KTM, MPL, VSF and pH. Then to make the framing of each sample *j* data the SLL value was calculated to matrix *U* that was limited to:

$$\sum_{h=1}^{5} u_{h,j} = 1 \tag{4}$$

The term $u_{h,j}$ is associated to safety level degree of each sample *j* data. The product of multiplication between $d_{h,j}$ and $u_{h,j}$ resulted in a value $D_{h,j}$ that was resolved by a objective function:

$$min\left\{F\left(u_{h,j}\right) = \sum_{h=1}^{5} D_{h,j}^{2}\right\} \tag{5}$$

Based on equation (4) to Lagrange function (L) was obtained to:

$$L(u_{h,j};\lambda_j) = \sum_{h=1}^{5} u_{h,j}^{2} d_{h,j}^{2} - \lambda_j \left(\sum_{h=1}^{5} u_{h,j} - 1\right) \tag{6}$$

Thus based on [7] was suggestion a equation to estimate the value of safety level degree, nominally SSL; for a sample *j* through $u_{h,j}$ equal:

$$u_{h,j} = \left(t_{h,j}^{2} \sum_{h=} t_{h,j}^{-}\right) \tag{7}$$

The condition is that $d_{h,j} \neq 0$. On order hand the resolution of equation (7) was made possible through of Lagrange multiple (λ_j). This way the relation between the partial derivate $\partial L/\partial u_{h,j}$ and $\partial L/\partial \lambda_j$ was equalled to value 0. However the measure of safety level degree of each sample *j* data couldn't be directly determined in the matrix form. Thus the a characteristic value to $u^{*}_{h,j}$ was calculated on axis similar to mass center like the mechanics when the center represent the level *h* and $u_{h,j}$ in coordinate plan of $h \sim u^{*}_{h,j}$ represented in matrix $U^{*} = (u^{*}_{h,j})$:

$$U^{*} = \begin{bmatrix} u^{*}_{1,1} & u^{*}_{1,2} & \dots & u^{*}_{1,38} \\ u^{*}_{2,1} & u^{*}_{2,2} & \dots & u^{*}_{2,38} \\ u^{*}_{3,1} & u^{*}_{3,2} & \dots & u^{*}_{3,38} \\ u^{*}_{4,1} & u^{*}_{4,2} & \dots & u^{*}_{4,38} \\ u^{*}_{5,1} & u^{*}_{5,2} & \dots & u^{*}_{5,38} \end{bmatrix} = u^{*}_{h,j}) \tag{8}$$

Thus it was possible to calculate the SSL_j value to each fuel station ($n = 38$) based on:

$$SSL_j = \left(1,2,3,4,5\right) u^{*}_{j} = \sum_{h=1}^{5} u^{*}_{j}.h \tag{9}$$

However the MPR reading will be done through of numerical value of SSL. The intervals those values will classify the safety level degree express of each fuel station available (Table 1).

4 Results a PSM application

This proposal was applied in area of Belém city of approximately 1.3 km², located in state of Pará, north region of Brazil. For this research was available a 102 wells data obtained in [6]. Thus it was possible obtained the NVI*f* value by FIS with support toolboxes on Matlab®. The NVI*f* were interpolation through ordinary kriging. The experimental semivariogram was better adjusted for SPH model resulting in SQR > 0 and $R^2 = 0.71$.

In an experimental area, 38 fuel stations were analysed. The simulation of susceptibility criteria was materialized in six scenarios (constant criteria: $5.5 \leq pH \leq 6.5$ and VSF equal to 2). Considerate this scenarios the SSL value of each fuel station were proportional to growing of categories. Almost all of fuel station showed between $5 \leq SLL > 4$ in the first scenario (Figure 1a) and $4 \leq SLL > 3$ in the second until the fifth scenario (Figure 1a – 1e). The last scenario all fuel stations showed values that qualified the security level as "W" (Figure 1f).

Finally this work showed that the application of model recognition pattern is a tool to management groundwater resources opposite the possible leakage of fuel. Therefore, this methodology that qualified a safety level of a fuel station considering the criteria of vulnerability and susceptibility is an important strategy to making decision in planning and monitoring of the fuel station in urban area.

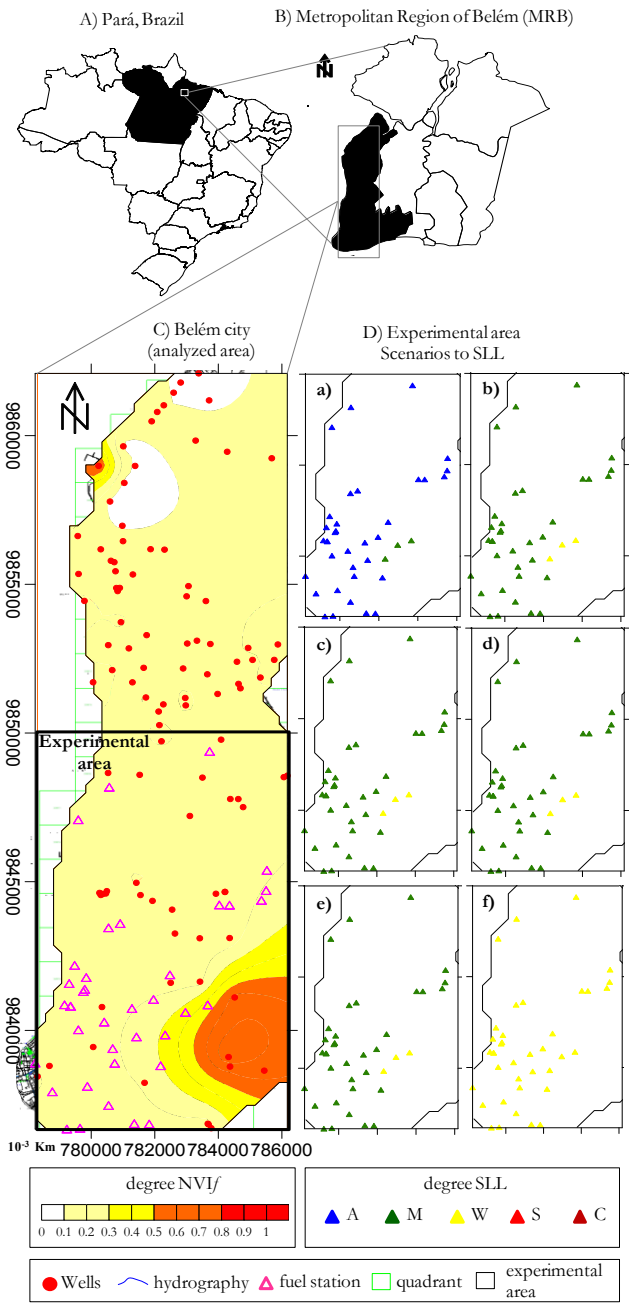


Figure 1: Area analysed to scenarios SSL: a) KTM = 1, MPL = 2 and TOP = 1; b) KTM = 1, MPL = 2 and TOP = 2; c) KTM = 2, MPL = 2 and TOP = 2; d) KTM = 3, MPL = 2 and TOP = 2; e) KTM = 3, MPL = 3 and TOP = 2 and f) KTM = 4, MPL = 3 and TOP = 3.

References

[1] J. Casana and G. Dixon. Application of quantitative risk assessment evaluation of underground storage tanks to insurance, banking, and real estate transactions. In *Petroleum Contaminated Soils* - V.2. E.J. Calabrese & P.T. Kostecki (eds.) Lewis Publishers, Michigan, USA, p. 321 – 327 (1990).

[2] S.S.D. Foster and R.C.A. Hirata. Groundwater pollution risk assessment; a methodology using available data. In *Groundwater pollution risk assessment; a methodology using available data. CEPIS* (1998);

[3] E.G. Lisboa; A.R.B. Barp; R.L. Mendes (2013) Fuzzy-Geostatistics Approach for Mapping the Intrinsic Vulnerability of Aquifers. *RBRH: Brazilian Magazine Water Resources*, Porto Alegre, v.18, n.1,111-123.

[4] S. Y. Chen (1996). Non-structured decision making analysis and fuzzy optimum seeking theory for multi-objective systems. *J. Fuzzy Math*. 4(2), 835–842

[5] T.L. Satty. Decision making for leaders: the analytic hierarchy process for decisions in a complex world, *RWS publications*. 1999.

[6] Geological Survey of Brazil-CPRM. Project of Hydrogeologic Studies of Metropolitan Region of Belém and adjacencies. Belém: *CPRM*, 2001. 110 p.

Optical Character Recognition of Guinea and Bissau boundaries documentation

Henrique Mata[1], João R. Caldas[1] Pinto and Maria C. S Nunes[2]
henrique.mata@tecnico.ulisboa.pt, joao.c.pinto@tecnico.ulisboa.utl.pt, mcarmonunes15@gmail.com

[1]IDMEC/ LAETA, Instituto Superior Técnico, Universidade de Lisboa

[2]IICT, Instituto de Investigação Científica Tropical, Travessa Conde da Ribeira, 9, B, 1º, Lisboa, Portugal

Abstract

This paper describes the image processing techniques that will be applied to an assembling of historical and diplomatic documents, with all the relevant information regarding the Guinea Bissau boundaries, in order to improve the performance of commercial Optical Character Recognition (OCR) software, when applied to degraded documents. An algorithm is proposed with encouraging results and may lead to a complete package of software that can be applied to organize documents relative to the others CPLP countries.

1 Introduction

Optical character recognition (OCR) is a practical application of state-of-the-art image processing and pattern recognition developments. The uses of OCR include digital document archiving, printed text search and automated form processing [1]. Nevertheless, while modern printed text can be recognized very accurately, with commercially available software, performing OCR on more exotic material (such as gothic fonts, ancient typesets and handwriting) is currently and noticeably less successful [2].

In this project we apply a set of image processing techniques to printed ancient documents in order to improve their optical character recognition (OCR) performance. Because it is difficult to develop a better OCR than the best ones currently available on the market, we chose to pick one of the best, ABBYY FineReader [3], and seek to restore the documents to increase the effectiveness of this package.

The original data will be submitted to an analogue-digital conversion, and image processing techniques will be applied to remove residual artifacts using morphological methods, combined with other image enhancement functions, improving in this way the documents appearance and reducing storage volume.

2 The problem

The main issues present in this analysis are related with the multiple signs of degradation present in the documents. Most of these signs result from the aging process, especially if poor care was taken to store and preserve the documents over time. The wide variety of documents also arise difficulties in finding a general approach to restore the set of documents (see fig.1).

3 Implemented Techniques

Resampling:

A factor responsible for Finereader performance losses is the presence of noise, acquired during the scanning process [6]. Thus, due to the loss of information that occurs in the resampling process, this arises as a possible solution to minimize the noise effect, increasing the Finereader's hit rate. Working with smaller images will also allow speeding up the analysis and reducing the storage memory needed. The reduction factors used in this study are 1,00; 0,75, 0,50 and 0,25.

Segmentation:

Image segmentation is a common first step to document image analysis, usually performed in the preprocessing stage of image processing applications such as optical character recognition (OCR) [2]. Through segmentation, it will be possible to distinguish different areas of interest within an image. This includes distinguishing the foreground from all background and interference.

Because the background interference usually presents intensity levels that are a mid-term between the background and the foreground, thresholding and clustering methods can be used to separate these defects from the rest of the image. However, there are also cases of extreme degradation in which more sophisticated methods have to be used [1]. The final result should generate a good quality monochromatic image responsible for increasing the OCR software hit rate.

Deblurring:

Noise and blurring of images can often not be avoided in certain imaging systems. This is due to the inherent properties of the imaging system itself. Hence to reduce or remove this blurring effect, algorithms that take the imaging process into consideration perform better. The efficacy of the deblurring algorithm depends on the accuracy with which the imaging process has been modeled. This is because the deblurring process is ill posed. Deblurring will be applied with the purpose of sharpening the foreground, in order to achieve a higher definition of the characters making the process of recognition to become more efficient.

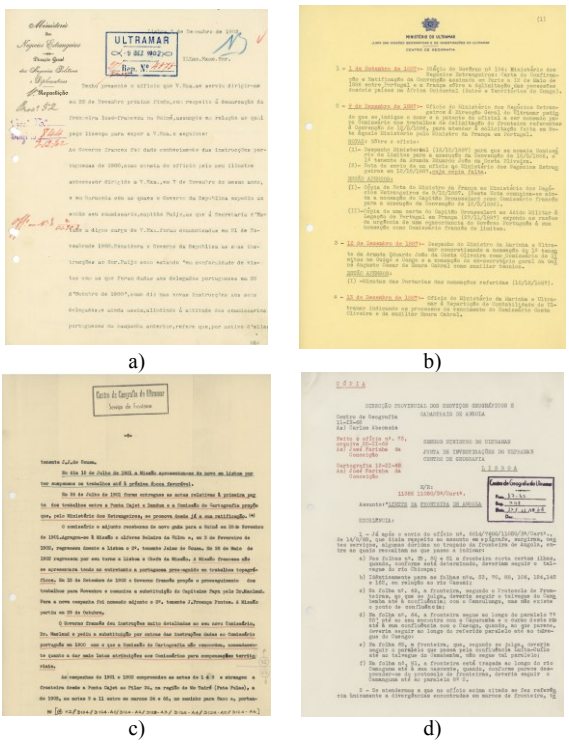


Fig. 1 – Examples of images of Guinea Bissau documents

4 Results

The goal of increasing Finereader's performance using resampling was achieved with the increase in the sensibility, percentage of characters correctly recognized and percentage of words correctly recognized, using a 0,75 reduction factor, for modern documents (see Table 1). The number of characters analysed was 6995 in 1440 words.

Table 1 – Results of Resize

%	Precision	Sensibility	Specificity	Characters C.R.*
1,00	100,00	96,04	100,00	95,79
0,75	99,97	96,43	89,58	96,18
0,50	99,99	95,83	95,65	95,59
0,25	98,24	91,93	50,51	88,96

*C.R. : Correctly Recognized

Concerning the segmentation techniques analysed (see Table 2), although improvements were only occurred in the specificity, this approach reduced drastically the size of the image, and speeded up the Finereader analysis, without affecting significantly the Finereader's performance. However, none of the methods could improve Finereader's precision or the percentage of characters correctly recognized. In this sense, the main gain obtained from this technique is related with the images size reduction.

Table 2 – Results of the Segmentations Methods

%	Precision	Sensibility	Specificity	Characters C.R.
Sauvola	98,60	93,67	54,97	90,99
Canny +Sauvola	98,85	93,37	60,99	90,83
Fuzzy	98,42	94,27	49,47	91,52
Otsu	97,38	88,50	60,21	83,56

*C.R. : Correctly Recognized

Regarding deblurring, none of the methods analysed showed any significant improvement. In addition significant losses in the statistical measures occurred (see Table 3). Concerning the size of the image, this method did not affect it significantly.

Table 3 – Results of the Deblurring Methods

%	Precision	Sensibility	Specificity	Characters C.R.
Wiener	77,91	70,25	53,67	49,13
CLS	84,52	71,66	84,52	54,73
Lucy	85,83	75,67	58,28	58,24
Blind	68,96	62,95	48,76	39,78

*C.R. : Correctly Recognized

5 Conclusions and future work

It was shown that resizing the image to a lower dimension has all the advantages of manipulating shorter files without degrading the OCR performance even improving some measures. This is true till a reduction factor of about 0,5. Other techniques were experimented to improve the OCR performance such us several adaptive binarization and deblurring algorithms. Although some improvements were registered results are not yet satisfactory when the documents ink is too faded.

Acknowledgments

This work is funded by National Founds through FCT - Foundation for Science and Technology (Fundação para a Ciência e Tecnologia) under the Project PTDC/ATP-GEO/4645/2012,IDMEC.

References

1. Castro, P., *Restauro de Documentos Antigos por Processamento de Imagem (Master's thesis)*. Instituto Superior Técnico, Lisboa, 2007.
2. J. M. C. Sousa, J. R. Caldas Pinto, C. S. Ribeiro and J. M. Gil, *Ancient document recognition using fuzzy methods*, Fuzz – IEEE 2005, 2005.
3. João R. Caldas Pinto, Lourenço Bandeira, João M. C. Sousa, Pedro Pina, "Combining Fuzzy Clustering and Morphological Methods for Old Documents Recovery" , 2nd Iberian Conference on Pattern Recognition and Image Analysis (IbPRIA05), June 7-9, 2005, Estoril, Portugal,.Lectures Notes in Computer Science, LNCS 3523, pp. 387–394, Springer-Verlag Berlin Heidelberg 2005.
4. F. Muge, I. Granado, M. Mengucci, P. Pina, V. Ramos, N. Sirakov, J.R. Caldas Pinto. "Automatic Feature Extraction and Recognition for Digital Access of Books of the Renaissance", ECDL 2000, pp. 1–13, Springer-Verlag Berlin Heidelberg 2000.
5. R.C. Gonzalez, R.E. Woods, S.L. Eddins. *Digital Image Processing Using MATLAB*. Gatesmark Publishing, 2009.
6. Finereader, Inc. http://www.finereader.abby.com

Toward Building an Automatic Video Surveillance System

Samaneh Khoshrou[12]
skhoshrou@inescporto.pt
Jaime S. Cardoso[12]
jaime.cardoso@inescporto.pt
Luís F. Teixeira[2]
luisft@fe.up.pt

[1] INESC TEC
Porto, Portugal
[2] Faculdade de Engenharia
Universidade do Porto

Abstract

Nowadays, video surveillance systems are taking the first steps toward automation, in order to ease the burden on human resources as well avoid human error. In this paper, we introduce a framework to classify evolving data streams generated in a multi-camera surveillance scenario applying a class-based ensemble. The framework exploits active learning, in order to interact wisely with operators to address various problems that exist in such non-stationary environments, such as concept drift and concept evolution. Favourable results indicate the success of the framework.

1 Introduction

Surveillance systems are spreading rapidly over the past few years, capturing 24/7 which provides massive amount of information. The evolving nature of captured data makes extraction of the meaningful information further complex. In such environments, it is more than likely that either due to intrinsic changes (pose change, movement, etc.), or extrinsic changes (lighting conditions, dynamic background, changes in camera angle, etc.), the distribution of data changes over time (concept drift). Moreover, new object(s) can enter the scene (class evolution), thus new models need to be trained.

Figure 1 demonstrates a typical surveillance scenario. When entering the scene, the object will enter the coverage area of at least one of the cameras. In such environments where objects move around and cross the FOV of multiple cameras, it is more than likely to have multiple streams, potentially overlapping in time, recorded at different starting points with various lengths, for the same individual object. Herein, we look at the problem as learning from multiple data streams (visual data) in wild environments, that views the whole or segments of a stream as a unique element to classify. Parallel stream mining methods in the literature [3] require equal-length streams coming from a fixed number of sources. Thus, they would fail to leverage information from time-varying video tracks. Learning from time-changing data has mostly appeared in a data mining context and various approaches have been proposed. Ensemble-based approaches constitute a widely popular group of these algorithms to handle concept drift [1] and in some recent works class evolution [4], as well. Ensembles provide a natural mechanism to update a knowledge by adding, removing or updating the knowledge. Most approaches use a fixed size ensemble in where a classifier is trained from the most recent knowledge. This classifier replaces the oldest or the least contributing member. Although discarding the classifier controls the complexity, it brings up some issues, because by discarding a classifier, we simply throw away the knowledge of all the classes available in the chunk. Thus, if a class appears after a long time, it may be misclassified as most of the classifiers may include the information of the most frequent and recent classes. This problem may be addressed with approaches where all the classifiers are kept in the ensemble. This work extends our prior framework, Never Ending Visual Information Learning (NEVIL) [5]. NEVIL exploits *Discriminative approaches* in order to actively classify parallel video streams. Since NEVIL computes the posterior probability (that must sum to 1), it is likely

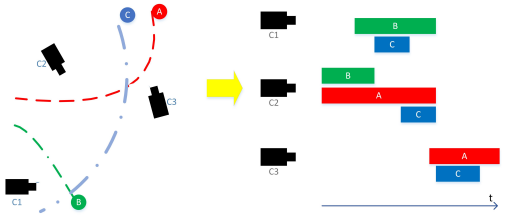


Figure 1: Typical surveillance scenario

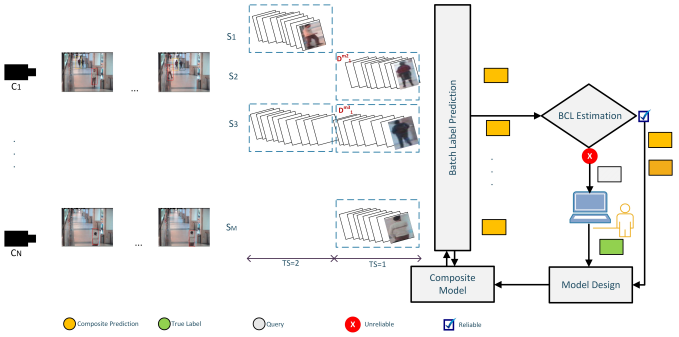


Figure 2: High-level overview of the framework

to assign a high enough (reliable) probability to a new class and mislead the system. Thus in order to get higher accuracy we need to spend more human resources. To address this problem, a class-based ensemble is introduced, where models of each class are stored separately. Class-based ensemble is firstly introduced in [2] where a model is trained for each class in a chunk. The ensemble keeps a fixed size ensemble of each class and it has been shown that this approach is more robust than traditional ensembles, though it needs the presence of training partition in each chunk. Although this approach does not fit directly in our scenario, we expect that class-based ensemble can improve our framework.

2 Learning algorithm

In this section we present NEVIL.g framework (see Fig. 2), that is designed for learning from non-stationary environments where no labelled data is available. The algorithm is a one-pass class-based ensemble of classifiers that trains a separate model (h_t^j) for a class (j) at every time slot (t). It also keeps models of each class in a separate ensemble (*Micro-Ensemble*). A time-adjusted weighting strategy combines the probabilities outputted by the models in order to make the final decision.

The framework receives multiple visual streams, generated by a typical tracking algorithm, which analyses sequential video frames and outputs the movement of targets between the frames. Environmental challenges such as varying illumination, lack of contrast, bad positioning of acquisition devices, blurring caused by motion as well as occlusion make data often noisy and/or partially missing. We address these challenges by a batch divisive strategy, as learning from a data batch may reduce the noise and fill the gaps caused by miss-tracking.

The algorithm is provided with a series of data batches $\mathcal{D}_t^{m_i}$, where m_i is the index of the i_{th} stream present at time slot t, TS_t, (not all streams are necessarily present). Note that a stream corresponds to a track generated by the tracking system and a single camera can yield multiple streams. A single batch aggregates B frames. The starting time of each stream is potentially different from stream to stream but batches are aligned between streams. Inside each frame the data corresponds to some pre-selected object representation (e.g. bag of words).

The ensemble obtained by all *Micro-Ensemble (ME)* generated up to the current time slot TS_t is named the composite hypothesis H_{t-1}. With the arrival of the current data batches $\mathcal{D}_t^{m_i}, \quad i = 1 \cdots M$, the algorithm tries to predict the most probable class label for each of the batches in current TS_t based on the probability estimate $p(C_k|\mathcal{D}_t^{m_i}, H_{t-1})$, where C_k runs over all the class labels observed so far.

This kind of on-line learning approach addressed in this work can suffer if labelling errors accumulate, which is inevitable. Unrelated objects will sooner or later be assigned the same label or different labels will be assigned to different views of the same object. To help mitigate this issue, we allow the system to interact with a human, to help it stay on track.

Initially, the composite model is initialized to yield the same probabil-

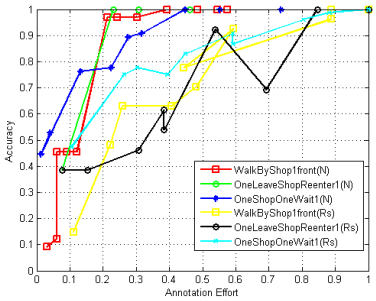


Figure 3: Comparison of the performance of Random Strategy and NEVIL.g on real scenarios

ity to every possible class (uniform prior). When the batches $\mathcal{D}_1^{m_t}$ in time slot t become available, the framework starts computing the probabilities $p(C_k|\mathcal{D}_t^{m_i}, H_{t-1})$ for each batch $\mathcal{D}_t^{m_i}$ in the time slot. Once $p(C_k|\mathcal{D}_t^{m_i}, H_{t-1})$ is obtained, a batch confidence label (BCL) is estimated; if BCL is high enough (above a predefined threshold), the predicted label

$$\arg\max_{C_k} p(C_k|\mathcal{D}_t^{m_i}, H_{t-1})$$

is accepted as correct; otherwise the user is requested to label the data batch. The labelled batches (either automatically or manually) are used to generate new separate models h_t^k (k runs over all the classes available in t) that are kept in micro ensembles ME_k, integrating in the composite, yielding H_t.

3 Experimental Methodology

3.1 Datasets

A series of experiments were conducted to explore the capabilities of the proposed framework. We tested the NEVIL framework with real video data from CAVIAR dataset [6] including: OneLeave ShopReenter1, Enter ExitCrossingPaths1, OneShopOneWait1, OneStop Enter2 and WalkBy Shop1front.These sequences present challenging situations with cluttered scenes, high rates of occlusion, different illumination conditions as well as different scales of the person being captured. We employ an automatic tracking approach to track objects in the scene and generate streams of bounding boxes, which define the tracked objets' positions. As the method may fail to perfectly track the targets, a stream often includes frames of distinct objects [7]. An hierarchical bag-of-visterms method is applied to represent the tracked objects, resulting in a descriptor vector of size 11110 for each frame.

3.2 Evaluation Criterion

Active learning aims to achieve high accuracy using as little annotation effort as possible. Thus, a trade-off between accuracy and proportion of labelled data can be considered as one of the most informative measures. Let N denotes the total number of batches, MC refers to misclassified batches, then the accuracy of the system in a given time slot is formulated as ratio of correctly classified batches to total number of batches. The total accuracy of a system over a period of time is derived from the mean accuracy of all the time slots. The *Annotation Effort* is formulated as ratio of manually labelled batches to the total number of batches. One expects that the accuracy increases with the increase of annotation effort.

3.3 Baseline Approach: Random Strategy

The Random strategy [8] labels the incoming batches randomly instead of wisely deciding which batches is more informative. Constrained by budget, batches are sent for annotation.

3.4 Results

Firstly, multiple tests were run to determine the optimal batch size for each dataset to be explored. The batch size was varied between 1% to 50% of the size of the longest stream available in each dataset. Experiments were repeated for 50 equally spaced values in that range. The optimal batch size varies and is influenced by the characteristics of the streams present in each dataset, range between 25 and 30.

Figure 3 presents the results on multiple CAVIAR clips, where various lengths and number of streams from different classes are present, and compares NEVIL.g with *Random Strategy*. We see that even at big budget Random strategy may fail due to selection of non or less informative batches (see *OneLeave ShopReenter1* plots). Figure 4 shows the results of NEVIL.g on multiple synthetic scenarios. To the best of our knowledge, there is no approach (except NEVIL) that can classify parallel data streams while interacting with an oracle (there are however clustering approaches which are not applicable in our scenario). Thus we compared this framework with the best setting of NEVIL (SVM-based NEVIL). Comparing to NEVIL, our method has improved. For half of the clips (including *OneStopEnter2, WalkBy Shop1front,OneLeave ShopReenter1*), we obtain over 90% accuracy with a manual labelling of 20% of batches. Although the interaction increase for other scenarios, the value is quite acceptable considering the complexity of the data (we need to annotate 25% of batches to gain 85% correct classification for *OneStopMoveEnter1*).

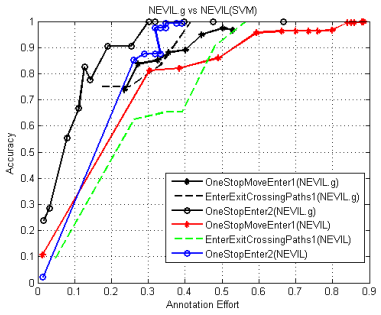


Figure 4: Performance evaluation on multiple CAVIAR clips.

4 Conclusions

We introduced a class-based ensemble framework for the classification of parallel visual data streams. The framework has shown promising performance with a fairly little human collaboration and can be applied in an on-line process. However, the growing complexity is our main concern as the method does not forget any knowledge. Hence, the main direction for future work would be controlling the complexity of the ensemble in order to make the framework applicable for never-ending scenarios.

Acknowledgment

This work is financed by the ERDF European Regional Development Fund through the COMPETE Programme (operational programme for competitiveness), FCOMP-01-0124-FEDER-037281 and by National Funds through the Fundação para a Ciência e a Tecnologia (FCT) (Portuguese Foundation for Science and Technology) within project PEst-C/EEI/LA0014/2013 and the grant SFRH/BD/80013/2011.

References

[1] Marcel R. Ackermann, Christiane Lammersen, Marcus Märtens, Christoph Raupach, Christian Sohler, and Kamil Swierkot. StreamKM++: A clustering algorithms for data streams. In *Journal of Experimental Algorithmics*, pages 173–187, 2010.

[2] Tahseen M. Al-Khateeb, Mohammad M. Masud, Latifur Khan, and Bhavani Thuraisingham. Cloud guided stream classification using class-based ensemble. In *Proceedings of the 2012 IEEE Fifth International Conference on Cloud Computing*, pages 694–701, 2012.

[3] Ling Chen, Lingjun Zou, and Li Tu. A clustering algorithm for multiple data streams based on spectral component similarity. *Information Sciences*, 183(1):35–47, 2012.

[4] Ryan Elwell and Robi Polikar. Incremental learning of concept drift in nonstationary environments. *IEEE Transactions on Neural Networks*, 22(10):1517–1531, 2011.

[5] Samaneh Khoshrou, Jaime S. Cardoso, and Luís Filipe Teixeira. Active mining of parallel video streams. *CoRR*, abs/1405.3382, 2014. URL `http://arxiv.org/abs/1405.3382`.

[6] CAVIAR project consortium. Caviar dataset, 2001. URL `http://homepages.inf.ed.ac.uk/rbf/CAVIAR/`.

[7] Luís Filipe Teixeira and Luís Corte-Real. Video object matching across multiple independent views using local descriptors and adaptive learning. *Pattern Recognition Letters*, 30(2):157–167, 2009.

[8] Indre Zliobaite, Albert Bifet, Bernhard Pfahringer, and Geoffrey Holmes. Active learning with drifting streaming data. *IEEE Trans. Neural Netw. Learning Syst.*, 25(1):27–39, 2014.

A rate-distortion study on microarray image compression

Luís M. O. Matos
luismatos@ua.pt
António J. R. Neves
an@ua.pt
Armando J. Pinho
ap@ua.pt

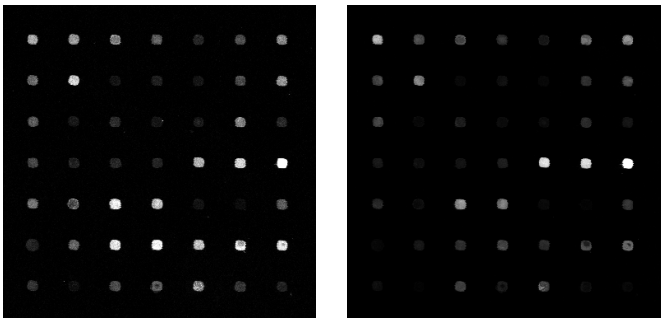
Signal Processing Laboratory
IEETA/DETI
University of Aveiro,
3810-193 Aveiro, Portugal

Abstract

In this paper we present a rate-distortion study on microarray image compression. We evaluate two image coding standards (JBIG and JPEG2000), a method based on bitplane decomposition, and a method recently proposed based on binary-tree decomposition. According to the results obtained JPEG2000 attains the worse compression results particularly for higher bitrates. On the other hand, the method based on binary-tree decomposition outperformed all the other compression methods in terms of L2-norm and L_∞-norm. The results attained for the binary-tree decomposition method can be explained due to its nature, consisting on coding each node of the tree with the goal of minimizing the L_∞ error.

1 Introduction

The raw data resulting from a microarray experiment consist of a pair of 16 bits per pixels of grayscale images (see Figure 1). These images, depend on the size of the array and the resolution of the scanner, and may require tens of megabytes to be stored or transmitted without any compression loss. Due to this fact, and the need for long-term storage and efficient transmission, the development of lossless compression methods with progressive decoding capabilities is an important challenge. Moreover, this progressive decoding capability is useful because the decoding process can be stopped at any time obtaining an intermediate image with some loss. This loss control can be based on the same principles that are used to extract genetic information from the microarrays.



Channel 1 Channel 2

Figure 1: An example of a microarray experiment (crop portion of "Def661" image from the ISREC dataset).

2 Related work and goals

In the literature we can find several lossless compression methods for microarrays images [1, 3, 4]. One interesting characteristic of some of these methods is its capability of progressive decoding the compressed image. This characteristic is very important because we can stop the decoding process at any time obtaining a partial image with some loss, loss that will not degrade the genetic information that will be extracted later. Besides that, the progressive decoding approach allows us to also decode the original image without any loss. This is also very important because the analysis techniques are subject to changes which means that in the future we may require to have the original image, without discarding any information.

Recently, we have proposed a lossless compression method for microarray images, based on a hierarchical organization of the intensity levels followed by finite-context modeling [2]. The organization of the intensity levels is attained by means of a binary tree. This binary tree is build and traversed in a way that each node that is expanded, is selected taking into account the smallest L_∞ reconstruction error.

Neves and Pinho proposed in 2009 a compression method based on a bitplane decomposition approach that uses a 3D finite-context model to drive the arithmetic coder [4]. Their work was inspired by EIDAC [5] but their method was improved by using image-dependent context models that are build with the goal of finding the "best" context configuration to encode each bitplane.

The goal of this research work is to evaluate in terms of rate-distortion this method based on a binary-tree decomposition [2], the method presented in 2009 by Neves and Pinho [4], and also two image coding standards (JBIG and JPEG2000) that have support for lossy-to-lossless compression.

3 Experimental results

The goal of this work is to evaluate the performance of four different methods in terms of rate-distortion. Two of the methods are specific for microarray compression [2, 4]. The other two methods are the JBIG and JPEG2000 standards.

In the literature, we can find several rate-distortion metrics that can be used to measure the performance of compression algorithms. In this research work we decided to use the root mean square error (or L2) and the maximum absolute error (L_∞). Because of the lack of space, we chose only two images to evaluate the performance of the methods mention earlier. One image of the Apo AI data set ("1230cG") and one image of the Microzip data set ("array1"). For the other data sets and images, the obtained results are quite similar so the conclusions made for these two images are the same as for the remaining images.

The rate-distortion curves of the two selected images can be found in Figs 2 and 3. The left charts are related to the L2-norm (root mean squared error) and the right ones were plotted in terms of L_∞-norm (maximum absolute error). Regarding the first metric (L2-norm), we can notice that lossless JPEG2000 provides similar rate-distortion results for lower bitrates (lower than 8 bpp). The poor results in higher bitrates for JPEG2000 could be explained by the default parameters used in JPEG2000. These default parameter values could not be suited for this kind of images with 16 bits per pixel and their particular characteristics in terms of noise, histogram, etc. Regarding the L_∞-norm, we observed that JPEG2000 has in fact the worst rate-distortion results. Similar to the previous metric, we can see that the rate-distortion results for JPEG2000 suffer a sudden deviation for higher bitrates which is probably related to the same problem pointed out earlier. When comparing methods [4] and [2], we can see that method [2] provides better rate-distortion results when compared to method [4]. The main reason to the previous statement is due to the nature of method [2]. As mentioned earlier in Section 2, method [2] has in its core a mechanism that minimizes the L_∞ error along the encoding/decoding process. This allows the method to obtain better rate-distortion results. Furthermore, method [4] was not designed with an error minimization goal in mind. It processes each bitplane of the microarray image without looking to any kind of error metric. The error after processing each bitplane depends on the remaning information that was not yet processed (on the lower bitplanes).

4 Conclusions

The goal of this study was to compare, in terms of rate-distortion, several compression methods. We used two image coding standards (JBIG and JPEG2000) and two different specialized methods for microarray images [2, 4]. According to the obtained rate-distortion results, we con-

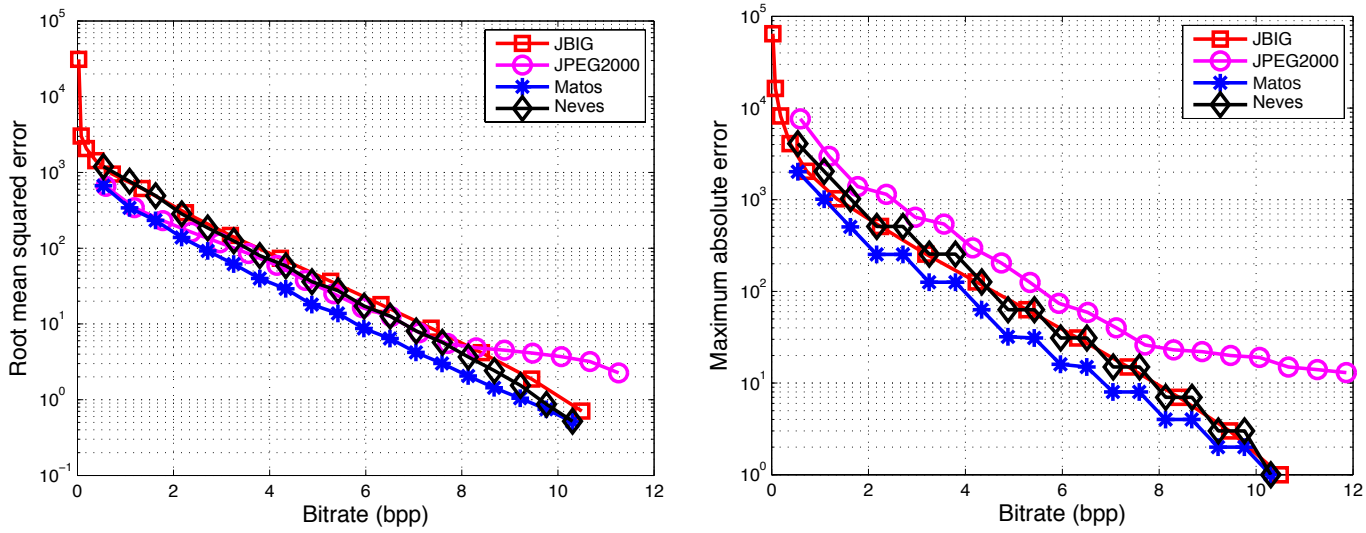


Figure 2: Rate-distortion curves for methods [2, 4], JBIG, and JPEG200, regarding image "1230c1G" of the Apo AI data set. Results are given in terms of L2-norm (root mean squared error) on the left chart and in terms of L_∞-norm (maximum absolute error) on the right chart. The curves indicated as "Matos" correspond to method [2], whereas the curves indicated as "Neves" correspond to method [4].

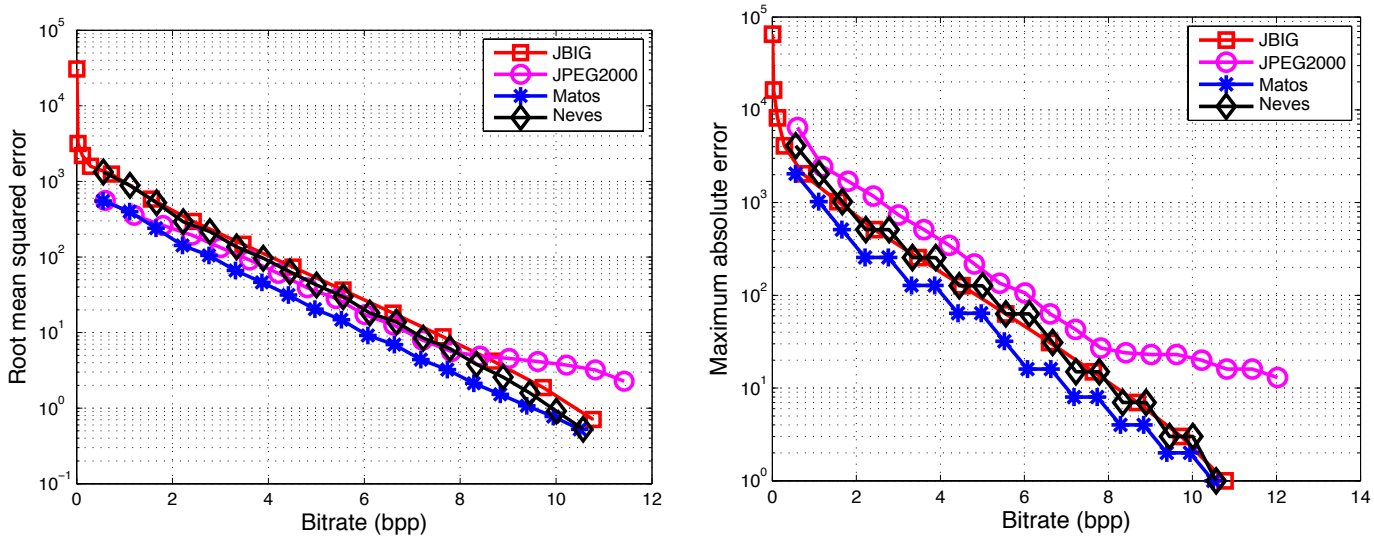


Figure 3: Rate-distortion curves for methods [2, 4], JBIG, and JPEG200, regarding image "array1" of the Microzip data set. Results are given in terms of L2-norm (root mean squared error) on the left chart and in terms of L_∞-norm (maximum absolute error) on the right chart. The curves indicated as "Matos" correspond to method [2], whereas the curves indicated as "Neves" correspond to method [4].

cluded that among the methods used in this work, JPEG2000 attained similar results to the other methods for lower bitrates (lower than 8 bpp), in terms of root mean squared error. The poor performance for higher bitrates is probably due to the default parameters used in the JPEG2000 codec. Regarding the other measure, the maximum absolute error, the JPEG2000 standard continues to attain the worse results when compared to the other method, mainly for higher bitrates. The phenomenon is probably caused by the same issue mention earlier. Regarding the specialized methods, the obtained results in both rate-distortion metrics are better when compared to JBIG and JPEG2000. In fact, the method based on binary-tree decomposition outperforms the other three methods due to its nature. The core method process each node of the tree in a way to minimize the L_∞ error.

Funding

This work was partially supported by FEDER through the Operational Program Competitiveness Factors - COMPETE and by National Funds through FCT - Foundation for Science and Technology, in the context of a PhD Grant (FCT reference SFRH/BD/86531/2012) and a project (FCT reference PEst-OE/EEI/UI0127/2014).

References

[1] S. Battiato and F. Rundo. A bio-inspired CNN with re-indexing engine for lossless DNA microarray compression and segmentation. In *Proc. of the IEEE Int. Conf. on Image Processing, ICIP-2009*, volume 1-6, pages 1737–1740, Cairo, Egypt, November 2009. doi: 10.1109/ICIP.2009.5413629.

[2] L. M. O. Matos, A. J. R. Neves, and A. J. Pinho. Compression of microarray images using a binary tree decomposition. In *Proceedings of the 22nd European Signal Processing Conference, EUSIPCO 2014*, September 2014.

[3] A. Neekabadi, S. Samavi, S. A. Razavi, N. Karimi, and S. Shirani. Lossless microarray image compression using region based predictors. In *Proc. of the IEEE Int. Conf. on Image Processing, ICIP-2007*, volume 2, pages 349–352, San Antonio, Texas, USA, September 2007.

[4] A. J. R. Neves and A. J. Pinho. Lossless compression of microarray images using image-dependent finite-context models. *IEEE Transactions on Medical Imaging*, 28(2):194–201, February 2009.

[5] Y. Yoo, Y. G. Kwon, and A. Ortega. Embedded image-domain compression using context models. In *Proceedings of the EEE Int. Conf. on Image Processing, ICIP-99*, volume I, pages 477–481, Kobe, Japan, October 1999.

Dynamic Sub-Sampling Line Scan Visual Inertial Odometry with Application to Mobile Robots

Ricardo Silva[1]
ricardolsilva@ua.pt
Vitor Santos[2]
vitor@ua.pt
Ricardo Pascoal[2]
rpascoal@ua.pt

[1] Department of Mechanical Engineering (DEM)
University of Aveiro, Campus Universitario de Santiago
3810-193 Aveiro, PT

[2] DEM & IEETA
University of Aveiro, Campus Universitario de Santiago
3810-193 Aveiro, PT

Abstract

This paper presents an odometry solution for automobiles which relies on the use of a line scan image sensor and an inertial sensor. Data from the image sensor was used to compute the velocity component which is parallel to line scan direction; the relative direction of the sensor was computed using data from two gyroscopes. An algorithm based on cross-correlation is proposed and applied to the line scan sensor. Several experiments were designed to test the method; the experiments took place indoors, under controlled conditions. These have enabled a characterization under imposed straight line movements, circular movements and movements with a variable distance between sensor and specimen. Based on the results, it is concluded that odometry with line scan sensors, aided by inertial sensors, has high potential for real world applications.

1 Introduction

Odometry is the process of measuring a robot's displacement. Most odometry solutions use encoders to measure wheel rotation, steering orientation and/or leg's movement, depending on the type of robot. Odometry based on analysing pulses from an encoder has the advantage that it can be self-contained, and it is always capable of providing an estimate of the vehicle position. The disadvantage of that type of odometry is that it must sum fractional displacement measurements to give position estimates, resulting in a method that is particularly sensitive to errors. Scale errors, such as those arising from incorrect modelling of wheel geometry, are the most common and may be due to fluctuation of tire pressure, wrong tire diameter, slippage of the wheels and tire tread wear.

Visual odometry is a technique that can estimate the image sensor motion using information from its own captured images; the position is defined as the incremental estimation of robot motion from image sequences using an on-board camera. Visual odometry is a method often used on off-road robots for planet surface exploration missions. The motion of the robot is estimated by calculating the pixel displacement between two consecutive frames [1]. Visual odometry has clear advantages over the common wheel's odometry. It is possible to measure velocity independent of the variations in tire pressure, tire diameter, uneven terrain and tire slip. It is possible to obtain a module which is compact and easy to install.

2 Related Work

Horn *et al*. [2] used a stereo camera looking forward to estimate the yaw and longitudinal velocity based on the relative movement of the surroundings. They used another camera pointed to the ground to estimate two dimensional velocity. The best results have been achieved with the camera facing the ground; the system was tested at 1 m/s. Chhaniyara *et al*. [3] used a matrix camera facing the ground to estimate the velocity; they experimented with the camera moving over sand and they obtained good results for low velocities, from 0 to 50 mm/s. Nourani-Vatani *et al*. [4, 5], obtained good results using a common webcam at 20 fps and an image size of 640 x 480 pixels; they used template matching with a searching area of 320 x 320 pixels. The acquisition speed was set to 20 fps due to the high processing time, 42 ms/frame, which is a clear limitation. Also, due to the low frame rate, the velocity of the vehicle was limited to 1.5 m/s, which is very low for common vehicles on the road. Nevertheless, this method provided better results than encoder based odometry.

The velocity range which can be sensed by the measuring unit is mostly influenced by the sampling rate, image size and resolution. If the velocity is too high, there will be no overlap of the frames and similarity between the successive images won't have the desired physical meaning, on the other hand, if velocity is very low then there may be no significant disparity between successive images; both situations result in incorrect estimated velocity. So, for car-like velocities it is necessary to employ a camera with very high sampling rate, which triggers challenges with respect to the necessary acquisition and processing capability. Kálmán [6] suggests the use of line scan cameras to achieve high velocities and resolution, using less computational resources than a common matrix camera. Through simulation, he has concluded that it is possible to measure movements of vehicles up to the range of 100m/s with high accuracy.

3 Hardware and Materials

To assess the proposed method, experiments were devised with several types of motion and precise ground truth. The movements were imposed using an industrial robot to ensure repeatability. To test the method with different textures, four samples of "ground" were used. With these samples it was possible to test the usability of the method in different texture sizes, regularity and color. The samples were concrete, stone, asphalt and cork. Illumination was controlled to ensure repeatability throughout the day. Sunlight has been blocked and artificial light from two 500 W halogen spots was used. The camera used on the experimental procedure was an industrial line scan camera of model P2-4x-04K40 (7 μm) from Teledyne DALSA. The optics were model Rodagon 50 mm f/2.8. The gyroscopes used are part of an Inertial Measuring Unit (IMU) of model POLOLU - MinIMU9DOF.

4 Experimental Procedure

The experimental setup is illustrated in Figure 1. To avoid vibrations and to ensure the repeatability of the experiments, the camera was kept static; the material samples were mounted on the robot and the movements were made in front of the camera. The trials were: pure straight line movements, both parallel and at an angle with the sensor to test the response under non-zero sideways motion; circular movements to simulate a vehicle turning; movements with variable focal distance to simulate the oscillations of the vehicle suspension. The total distance of the movements was 500 mm.

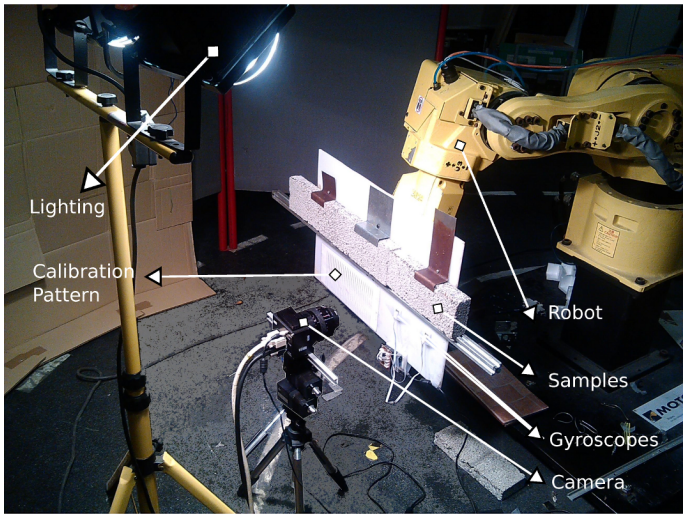


Figure 1: Illustration of the experimental setup.

5 Algorithms

To calculate the displacement between two line scans the Cross Correlation Theorem was invoked and the Fast Fourier Transform (FFT) used in discrete wave number space. The simplest algorithm would be to simply calculate the displacement between successive lines. This procedure was used but the results were not satisfactory for low velocities because the smallest detectable displacement was not achieved and subpixel resolution was insufficient. To solve this problem, the sensor and texture would need to be matched (increase resolution, lower frame rate or larger grain); another alternative is to reject intermediate lines. The improved algorithm consists in using one reference line, then, each captured line is correlated to the reference until a minimum displacement is found within a given limit. This technique is a form of temporal dynamic sub-sampling. After the minimum displacement is found, the current line becomes the new reference. The distance is calculated using cross-correlation; other distance metrics may be used. The results presented here were achieved using this improved algorithm.

6 Results

The optics magnification was kept fixed throughout the experiments; the field of view was 50.67 *mm* except where noted otherwise. The results presented here are all related to the experiments made with asphalt; it is important to point out that the results for the other materials were identical.

The results for the straight line movements are both shown in Table 1. The parallel one is represented as a movement at zero degrees and the one with non-zero orthogonal component is represented with 2.5, 5 and 10 degrees. The table shows the values of relative error for the average velocity and total traveled distance. The errors were calculated taking as reference the measurements of the robot. In Table 1 it is possible to note that the average velocity errors are roughly equal, this is due to the finite resolution of the sensor. The similarity of the results demonstrates that the method works consistently despite the non-zero othogonal component.

Table 1: Parallel movement - Summary of error (%) results using asphalt.

Angle (°)	Tgt. Vel. (mm/s)	Avg. Vel. (%)	Trvld. Dist. (%)
0	100	1.03	0.95
	500	1.03	0.99
	1000	1.03	2.23
2.5	100	1.02	0.79
	500	1.03	0.62
	1000	1.03	1.42
5	100	1.02	0.13
	500	1.03	0.33
	1000	1.03	1.24
10	100	0.86	1.10
	500	1.00	0.75
	1000	1.07	1.95

Figure 2 shows the typical plot for estimated and ground truth velocity of any of the circular motion experiments; Figure 3 shows the respective position plot. The two figures relate to a movement at 500 mm/s and 10 m diameter.

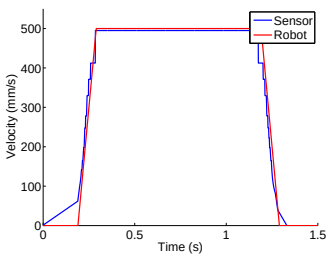


Figure 2: Velocity for circular movement: 500 mm/s, ø 10 m.

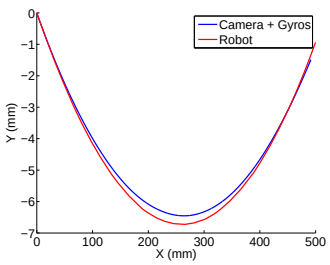


Figure 3: Position for circular movement: 500 mm/s, ø 10 m.

Resulting estimates for 2 m turning diameter were far from the ground truth. These results lead to the conclusion that it is possible to accurately measure velocities during turning diameters as tight as 5 m, however at 2 m the successive line scans start to lose physically interesting similarity and the correlation fails to provide reasonable estimates.

Changes in the field of view introduces additional error, as shown in Table 3, but were still under 3% for a motion amplitude of 25 mm. Further improvements and extended applicability may be achieved with a telecentric lens, as their magnification is invariable with distance.

Table 2: Summary of errors (%) for circular movement over asphalt.

Diameter (m)	Target Vel. (mm/s)	Average Vel.(%)	Trvld. Dist. (%)	Euclidean Dist. (mm)
2	100	6.13	5.30	36.44
	500	13.70	12.58	66.04
	1000	27.62	29.53	156.34
5	100	1.03	2.57	15.14
	500	1.03	0.71	7.35
	1000	1.03	2.27	17.70
10	100	1.03	0.13	6.09
	500	1.03	1.24	8.36
	1000	1.03	2.68	17.70

Table 3: Summary of errors for oscillatory movement over asphalt.

Target Velocity (mm/s)	Average Velocity Error (%)	Traveled Distance Error (%)
100	2.83	1.42
500	2.55	1.16
1000	2.65	1.26

7 Conclusion

It was shown that a line scan image sensor can operate as a velocity sensor under controlled conditions of lighting and in the absence of vibration. Under these conditions the sensor can estimate velocity with low error, about 1% in most of the cases. The procedure performed well for straight line motion with the camera aligned and misaligned up to a maximum of 10°. For circular motion performance was also interesting and provides confidence that it is applicable when an automobile maneuvers with a tight turning radius. For situations where the focal distance is changing the results were also satisfactory.

The method shows promising results and further investigation and development should be performed in order to study the applicability in real world scenarios.

References

[1] L. Matthies, Y. Cheng, M. W. Maimone. Visual Odometry on the Mars Exploration Rovers. *IEEE Robotics & Automation Magazine*, pages 54-62, 2006.

[2] Horn, Jan and Bachmann, Alexander and Dang, Thao. A Fusion Approach for Image-Based Measurement of Speed Over Ground. *2006 IEEE International Conference on Multisensor Fusion and Integration for Intelligent Systems*, pages 261-266, 2006.

[3] Chhaniyara, Savan. Optical flow algorithm for velocity estimation of ground vehicles: A feasibility study. *International Journal on Smart Sensing and Intelligent Systems*, 2008.

[4] Nourani-Vatani, N. and Roberts, J. and Srinivasan, M.V. Practical visual odometry for car-like vehicles. *IEEE International Conference on Robotics and Automation*, pages 3551-3557, 2009.

[5] Nourani-Vatani, N and Roberts, Jonathan and Srinivasan. IMU aided 3D visual odometry for car-like vehicles. *Australian Robotics and Automation Association*, 2008.

[6] Kálmán, V. On modeling and control of omnidirectional wheels. PhD thesis, *Budapest University of Technology and Economics*, 2013.

The authors would like to thank the FCT, Portuguese Foundation for Science and Technology, for partially funding the work through PEst-OE/EEI/UI0127/2014. The work of the third author has been financed through scholarship SFRH/BPD/69555/2010. The authors would also like to thank Infaimon for having lent the camera and framegrabber and Motofil for providing the robot.

RetinaCAD - Retinal Computer-Aided Diagnosis System

Behdad Dashtbozorg[1,2]
behdad.dashtbozorg@fe.up.pt

Ana Maria Mendonça[1,2]
amendon@fe.up.pt

Susana Penas[3,4]
susanaapenas@gmail.com

Jorge Polónia[3]
jjpolonia@gmail.com

Aurélio Campilho[2,5]
campilho@fe.up.pt

[1]INEB - Instituto de Engenharia Biomédica, Porto, Portugal

[2]Faculdade de Engenharia, Universidade do Porto, Porto, Portugal

[3]Faculdade de Medicina, Universidade do Porto, Porto, Portugal

[4]Centro Hospitalar São João, Porto, Portugal

[5]INESC TEC - INESC Science and Technology, Porto, Portugal

Abstract

This paper presents an automatic application that provides several retinal image analysis functionalities, namely vessel segmentation, vessel width estimation, artery/vein classification and optic disc segmentation. A pipeline of these methods allows the computation of important vessel related indexes, namely the Central Retinal Arteriolar Equivalent (CRAE), Central Retinal Venular Equivalent (CRVE) and Arteriolar-to-Venular Ratio (AVR), as well as various geometrical features associated with vessel bifurcations.

1 Introduction

Among several features associated with vascular changes, the Central Retinal Arteriolar Equivalent (CRAE), Central Retinal Venular Equivalent (CRVE) and Arteriolar-to-Venular Ratio (AVR) have been frequently used as indicators for the early detection, diagnosis, staging and follow-up of some systemic diseases, namely diabetes, hypertension, and vascular disorders, since they can reflect the narrowing or dilation of the retinal blood vessels caused by these diseases [4, 9]. These diseases also change the pattern of vessel branching, thus making the measurement of bifurcation geometrical features a useful tool for the diagnosis or prediction of such pathologic conditions [1].

In this paper, we introduce the RetinaCAD System, which is an automated application for a fast and reliable measurement of CRAE, CRVE, AVR values, as well as several geometrical features of the retinal vasculature. The evaluation of the system on images of a dataset from a local hospital shows a low failure rate and a significant correlation between the values calculated for distinct images of both eyes of the same patient.

2 RetinaCAD System

The RetinaCAD is a fully automatic system for the segmentation and classification of retinal structures and for the measurement of vascular features. This system can analyse optic disc centered retinal images with variable resolution and camera field of view (FOV). Fig. 1 shows the main interface of this application. In this section, we briefly describe the main tools for the detection and classification of retinal structures that were implemented in the RetinaCAD system.

2.1 RetinaCAD Image Analysis Tools

The main retinal image analysis tools provided by this application are: 1) vessel segmentation; 2) optic disc localization; 3) optic disc segmentation; 4) region of interest determination; 5) graph representation of the vascular tree; 6) A/V vessel classification.

1) Vessel Segmentation Tool: This tool is used for finding the vessels, and is the first stage for vessel caliber estimation and initial optic disc (OD) localization. The methods described in [7, 8] are applied for vessel segmentation. Fig. 2(b) illustrates the vascular tree which is generated by the tool for the original image of Fig. 2(a). Vessel caliber measurement

This work was supported by FEDER funds through the Programa Operacional Factores de Competitividade-COMPETE and by Portuguese funds through FCT-Fundação para a Ciência e a Tecnologia in the framework of the project PEst-C/SAU/LA0002/2013 and the research grant SFRH /BD/73376/2010.

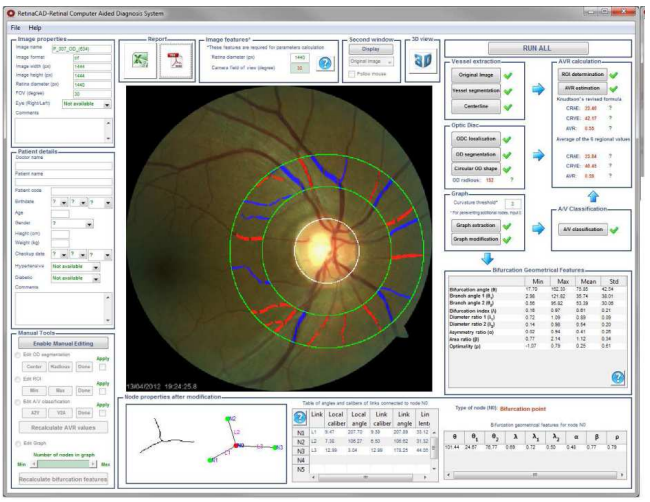


Figure 1: RetinaCAD interface.

is based on the distance transform of the segmented vascular tree; the result of this transform in each vessel pixel is its distance to the closest boundary point, d. After that, for each vessel centerline pixel, the vessel caliber value, vc, is estimated by $vc = 2d - 1$.

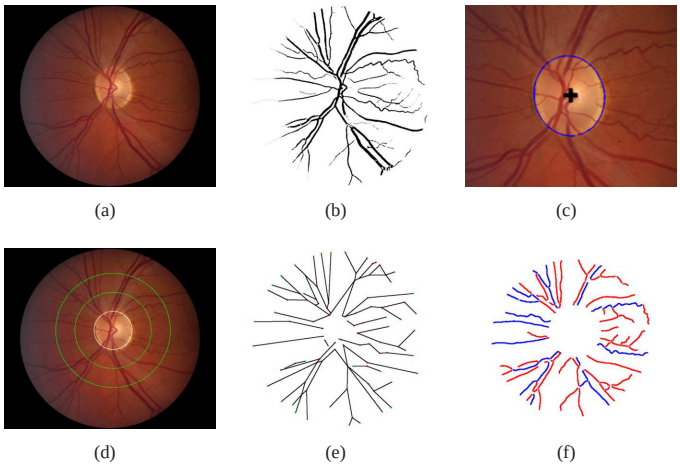
2) Optic Disc Localization Tool: The initial optic disc (OD) location is obtained following the approach based on the entropy of vascular directions described in [6] (Fig. 2(c)).

3) Optic Disc Segmentation Tool: A multiresolution sliding band filter (SBF), centered in the initial OD location is used for segmenting the OD boundary [2] (Fig. 2(c)). The final optic disc center (ODC) and OD radius are found by fitting a circle to the extracted boundary (Fig. 2(d)).

4) Region of Interest Determination Tool: The CRAE, CRVE and AVR values are calculated from the calibers of the vessels inside a specific region of interest (ROI), defined as the standard ring sector centered on the ODC, within 2 to 3 disc radius from the OD margin [4]. The corresponding ROI for CRAE, CRVE and AVR calculation is shown in Fig. 2(d).

5) Graph Representation Tool: The vessels centerline image is obtained by applying an iterative thinning algorithm to the vessel segmentation result. Then the graph nodes are extracted from the centerline image by finding the intersection points, the bifurcation points, the high-curvature points and the terminal points . After that each vessel segment is represented by a link between two nodes (Fig. 2(e)). The graph representation tool can display local and global features of the links (vessel segments) and nodes (bifurcation/intersection points), namely the caliber and the angle between vessel segments as well as the type of intersection (crossing or bifurcation).

6) Artery/Vein Classification Tool: This tool is based on the graph representation of the retinal vasculature as described in [3]. This method classifies the entire vascular tree by deciding on the type of each vessel intersection point (graph nodes) and by assigning one of two labels to each vessel segment (graph links). Final classification of a vessel segment as an artery or a vein is performed through the combination of the graph-based labeling results with intensity information from original color image. The result of A/V classification for the whole image is shown in Fig. 2(f).



(a) (b) (c) (d) (e) (f)

Figure 2: (a) Input image; (b) Vessel segmentation result; (c) Initial OD location (black cross) and OD boundary (blue); (d) ROI (delimited by the two green circles) and circular OD margin (white circle); (e) Graph representation of vascular tree; (f) A/V classification result.

Table 1: Comparison of Mean ± standard deviation (SD) and Pearson correlation between images of the right and left eyes.

Measurements	RetinaCAD				BDES [10]	Leung [5]
	Mean ± SD		Mean error±SD	Corr.	Corr.	Corr.
	Right eye	Left eye				
CRAE (μm)	158.1±15.9	161.9±16.9	11.2±8.5	0.66	0.71	0.70
CRVE (μm)	252.3±24.4	255.9±24.9	16.1±12.6	0.67	0.74	0.77
AVR	0.63±0.09	0.64±0.08	0.06±0.05	0.54	0.49	0.54

Corr.: Pearson's correlation coefficient.

2.2 Measurements

The CRAE and CRVE values are calculated based on Knudtson's revised formulas [4]. The AVR is defined as the quotient between CRAE and CRVE. For the AVR calculation, the ROI is equidistantly sampled to provide six regions. For each region, the six largest arteries and the six largest veins are identified, and the regional AVR is calculated. The final AVR estimate for the complete image is the average of the six regional values.

The branching angles are the basic measurements related to a bifurcation, and these features are obtained from the graph representing the vascular tree, as mentioned before. The branching angles and vessel segment calibers are used to derive features, such as the bifurcation index, asymmetry ratio, diameter ratio, area ratio and optimality [1].

3 Results

The system was evaluated using the images of a dataset from Centro Hospitalar São João (CHSJ dataset). This dataset contains 564 images from 141 subjects, where for each subject there are four images, two from the right eye and two from the left eye; the two images of each eye were acquired with two different FOV values (45° and 30°). The images of this dataset were used for investigating the robustness of the system to the use of distinct images of the same subject, through the assessment of the correlation between measurements from images of the right and left eyes and from images of same eye with different FOV. For the 564 images that were analysed using RetinaCAD, the results for 11 subjects were not accepted as a consequence of errors in A/V classification or OD segmentation. Although these errors can be solved using the manual modification tool which is included in this application, all the images of the subjects where the automatic procedures have failed were excluded. In the following, we discuss the results obtained for 520 images from 130 subjects.

The mean error and the Pearson correlation coefficient between measurements from the right and left eyes in the CHSJ dataset are shown in Table 1, where the correlation coefficients reported by the Beaver Dam Eye Study (BDES) [10] and Leung *et al.* [5] are also included. These results show a good correlation between right and left eyes for CRAE and CRVE and a moderate correlation for AVR, being similar to those reported in [10] and [5]. In order to evaluate the robustness of the methods in RetinaCAD, we have compared the results of the CRAE, CRVE and AVR values in images of the same eye that were acquired with a different FOV (45° and 30°). As can be observed in Table 2, all indicators show a small mean error and there is a significant correlation between images of the same eye that were acquired in distinct conditions, thus indicating the good performance and consistency of RetinaCAD.

Table 2: Comparison of results between images of the same eye with 45° and 30° FOV.

Measurements	Mean ± SD		Mean error±SD	Correlation coefficient
	FOV: 45°	FOV: 30°		
CRAE (μm)	154.7±17.5	162.9±16.4	11.4±8.8	0.76
CRVE (μm)	245.6±25.1	260.9±23.7	18.8±12.3	0.80
AVR	0.64±0.08	0.63±0.08	0.05±0.03	0.73

4 Conclusions

We have developed a user-friendly system, RetinaCAD, that is able to automatically detect, measure and classify two main retinal landmarks, the optic disc and the vessels. RetinaCAD can measure several vascular features that are recognized as indicators for some prevalent systemic diseases. We have demonstrated that the correlation between right and left eyes was good for the CRAE and CRVE values, which suggests that the measurements from one eye can provide adequate information about the changes in vessel calibers. After comparing the measured values for images of the same eye but with different FOV, a significant correlation and a low mean error were achieved, thus allowing the conclusion that RetinaCAD is adequate both for research and for general clinical use.

References

[1] B. Al-Diri, A. Hunter, D. Steel, and M. Habib. Manual measurement of retinal bifurcation features. In *in Proc. Engineering in Medicine and Biology Society (EMBC)*, pages 4760–4764. January 2010.

[2] B. Dashtbozorg, A.M. Mendonça, and A. Campilho. Optic disc segmentation using the sliding band filter. *Computers in Biology and Medicine*, (submitted).

[3] B. Dashtbozorg, A. M. Mendonça, and A. Campilho. An automatic graph-based approach for artery/vein classification in retinal images. *IEEE Trans. on Imag. Process.*, 23(3):1073–1083, March 2014.

[4] M.D. Knudtson, K.E. Lee, L.D. Hubbard, T.Y. Wong, R. Klein, and B.E.K. Klein. Revised formulas for summarizing retinal vessel diameters. *Current eye research*, 27:143–149, 2003.

[5] H. Leung, M. Harry, J.J. Wang, E. Rochtchina, A.G. Tan, T.Y. Wong, L.D. Hubbard, R. Klein, and P. Mitchell. Clinical and epidemiological computer-assisted retinal vessel measurement in an older population: correlation between right and left eyes. *Clinical and Experimental Ophthalmology*, pages 326–330, 2003.

[6] A. Mendonça, F. Cardoso, A. Sousa, and A. Campilho. Automatic localization of the optic disc in retinal images based on the entropy of vascular directions. In *Image Analysis and Recognition*, volume 7325 of *Lecture Notes Comput. Sci.*, pages 424–431. 2012.

[7] A.M. Mendonça and A. Campilho. Segmentation of retinal blood vessels by combining the detection of centerlines and morphological reconstruction. *IEEE Trans. Med. Imag.*, 25:1200–1213, 2006.

[8] A.M. Mendonça, B. Dashtbozorg, and A. Campilho. Segmentation of the vascular network of the retina. In E. Ng, U. Rajendra Acharya, Aurelio Campilho, and Jasjit S. Suri, editors, *Image Analysis and Modeling in Opthalmology*, pages 85–109. CRC Press, 2014.

[9] N. Patton, T. Aslam, T. MacGillivray, A. Pattie, I. J. Deary, and B. Dhillon. Retinal vascular image analysis as a potential screening tool for cerebrovascular disease: a rationale based on homology between cerebral and retinal microvasculatures. *Journal of anatomy*, 206(4):319–348, 2005.

[10] T.Y. Wong, M.D. Knudtson, R. Klein, E.K.B. Klein, S.M. Meuer, and L.D. Hubbard. Computer-assisted measurement of retinal vessel diameters in the Beaver Dam Eye Study: methodology, correlation between eyes, and effect of refractive errors. *Ophthalmology*, 111 (6):1183–90, June 2004.

Using Support Vector Machine model for fault detection along a water canal

José Duarte[1]
d10401@alunos.uevora.pt
Luís Rato[1]
lmr@uevora.pt
Manuel Rijo[2]
rijo@uevora.pt

[1] Departamento de Informática, Escola de Ciências e Tecnologia, Centro de Inovação em Tecnologias da Informação, Universidade de Évora, Rua Romão Ramalho, 59, 7000-671 Évora, Portugal

[2] Departamento de Engenharia Rural, Escola de Ciências e Tecnologia, Núcleo de Hidráulica e Controlo de Canais, Universidade de Évora, Núcleo da Mitra, Apartado 94, 7002-554 Évora, Portugal

Abstract

This paper reports a work in progress, the training of a Support Vector Machine model to detect faults in an experimental water supply canal. The work took place at the experimental canal of Núcleo de Hidráulica e Controlo de Canais at the Universidade de Évora. The main objective is to identify faults in the water depth sensors and to detect unauthorized water withdrawals using pattern recognition. The preliminary accuracy tests, in and out of sample, have shown an accuracy over 90% to identify 28 different patterns.

1 Introduction

Fault Detection and Isolation (FDI) is a research field where knowlegde--based models have been used with some success. Artificial Neural Networks and Fuzzy Systems are often used but its dependency of large amounts of training data and the slow convergence speed leads them to lose ground to models mathematical-based [13].

Support Vector Machines (SVMs) are been used as classification tool with a huge success in research areas like computer vision [7], health [9] and entertainment [4]. Lately, many studies propose the use of SVM in FDI problems [2, 12, 13].

This paper reports the work, still in progress, of the creation of FDI model for an experimental water supply canal using a SVM.

1.1 Presentation of the water canal

The work took place at the experimental canal of Núcleo de Hidráulica e Controlo de Canais (NuHCC) [10] at the Universidade de Évora.It is a canal with 145m of length and it is divided in four pools. Each one separated by an undershot gate with an overshot gate at the canal end, as can be seen in Figure 1. Sensors installed at the upstream, center and downstream allows to monitor the water depth in each pool. The maximum depth is 900mm, is the equivalent to the height of the pools. There is an offtake valve at upstream of every gates that allows to implement water withdrawals. An electric MONOVAR valve controls the canal inlet with a maximum design flow of $0.09 m^3 s^{-1}$.

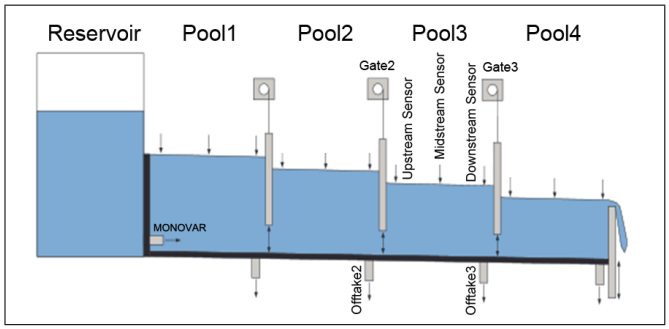


Figure 1: Diagram of the experimental water canal (adapted from[8])

The NuHCC facility is monitored and controlled by a MODBUS/Serial network of six Programmable Logic Controllers (PLCs) and a Supervisory Control and Data Acquisition (SCADA) system. A multi-platform controller interface [3] was used to interact with SCADA and collect data. Further details about this canal can be read in [8].

1.2 Objectives

At this moment, the focus is only in the third pool of the NuHCC canal. In Figure 2, it is possible to identify, from upstream to downstream, all the elements of interest to the study. Two offtakes in the left, two gates and water stream in the center and in the right the three water depth sensors.

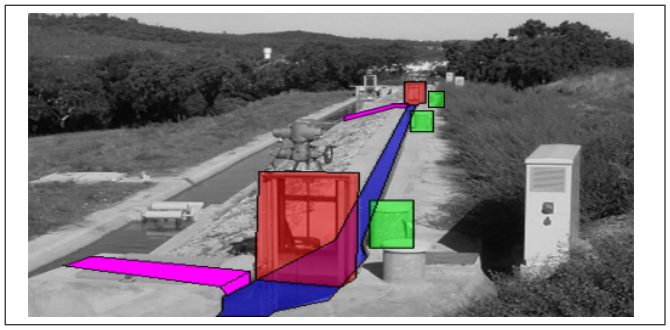


Figure 2: The third pool of the canal (Red - Gates 2 and 3; Pink - Offtakes 2 and 3; Green - Sensores Up, Center and Downstream).

The water depth sensors measure values between 0 and 900mm with an reading error of 0.005mm. The opening values of the offtake are measured in percentage, ie. from 0% to 100%. The height of the three first gates varies from 0 to 800mm, the last one from 0 to 700mm.

The main objective is to identify faults in the water depth sensors and to detect unauthorized water withdrawals using pattern recognition.

2 Preparing the data

The SVM is a popular and powerful classification technique. Considered easier to understand than Neural Networks, but users less familiar with it often have problems to get satisfactory results [6]. Classify the existing data is often the first task to create a model using a SVM. In this case. due to the nature of the problem, the first step was to collect the data.

Tests were made in order to collect the more significant data as possible. Tests like filling and emptying the pool were made and all readings of sensors and actuator recorded. The two offtake valves were used to simulate water uptake upstream or downstream of the pool. The sample time used was 1s.

2.1 Supervising the samples

The model should be able to classify correctly any given sample. To allow that to happen, every instance in the training set should be correctly classified. The model it is able to identify 28 different canal states. This means that there are at least 28 recognizable patterns (Table 1) provided by the training set.

2.2 Selecting the features

One problem detected in these kind of water depth sensors is inconsistent readings. It is an oscillatory behavior between the real value and a shift of it. Though temporary, this kind of error can lead of misclassifications. A threshold of 10mm has been defined in order to identify this error state. An instance of the training set that violates this condition it is classified as faulty sample. The resulting 26 labels are described in Table 1.

# of classes	Description
One	Nominal state
One	Loss of water
Three	Positive shift in one sensor
Three	Negative shift in one sensor
Three	Positive shift in two sensors
Three	Negative shift in two sensors
Six	Alternated shifts in two sensors
One	Positive shifts in three sensors
One	Negative shifts in three sensors
Six	Alternated shifts in three sensors

Table 1: Description of all classes

Selecting the right features for a SVM it is not a easy task. In fact, many studies were made about that matter [1, 5, 11]. The used features were kept to a minimum without compromising the performance. Just six features are used, the reading of the three depth sensors present in the pool and the difference between the actual and the previous reading. A further look into this difference (1) it is possible to understand it as the slope of the reading values along the time , for a sample time of $1s$.

$$\Delta sensor_i = \frac{sensor_i - sensor_{(i-1)}}{t_i - t_{(i-1)}} \quad (1)$$

An instance of the training set it is described in (2).

$$\overbrace{class}^{\text{Label}} \quad \overbrace{sensor_1 \;\; sensor_2 \;\; sensor_3 \;\; \Delta sensor_1 \;\; \Delta sensor_2 \;\; \Delta sensor_3}^{\text{Six features}} \quad (2)$$

3 Training and testing the model

Experimental tests were made in the canal in order to cover several nominal and faulty states. Seven data independent sets were created. One was used to train the model, the six others to test it. The nominal values of the test sets can been seen in Table 2. A linear kernel with the penalty parameter of the error term equal to 1 (ie. $C = 1$) was used.

Test #	MONOVAR (l/s)	Gates (mm)	Water level (mm)	Offtakes $(\%)$
1	$25 \rightarrow 25$	$0 \rightarrow 0$ (G3)	$0 \nearrow 800$	none
2	$25 \rightarrow 25$	$0 \rightarrow 0$ (G3)	$0 \nearrow 800$	none
3	$25 \searrow 0$	$400 \rightarrow 400$ (G4)	$400 \searrow 0$	$0 \nearrow 50$ (OT2)
4	$25 \searrow 0$	$400 \rightarrow 400$ (G4)	$400 \searrow 0$	$0 \nearrow 50$ (OT3)
5	$25 \searrow 0$	$400 \searrow 0$ (G4)	$400 \searrow 0$	$0 \nearrow 50$ (OT2)
6	$25 \searrow 0$	$400 \searrow 0$ (G4)	$400 \searrow 0$	$0 \nearrow 50$ (OT3)

Table 2: Nominal values for the six test sets

3.1 Faults detected and results

Three kinds of tests were made to measure the model accuracy. *In sample* test where the same set is used to train and test the model. The *10-Fold Cross-validation* technique has been used to estimate accuracy of the model. And then, six test sets were used to measure the *out of sample* accuracy. All tests have shown an accuracy over 90%. The results are presented in Table 3.

In sample accuracy
Accuracy = 100% (1038/1038) (classification)
Out of sample accuracy
Train set 1 - Accuracy = 99.3506% (918/924) (classification)
Train set 2 - Accuracy = 99.3506% (918/924) (classification)
Train set 3 - Accuracy = 98.1722% (1665/1696) (classification)
Train set 4 - Accuracy = 99.0538% (2303/2325) (classification)
Train set 5 - Accuracy = 90.8019% (385/424) (classification)
Train set 6 - Accuracy = 90.0463% (389/432) (classification)
All sets - Accuracy = 97.8141% (6578/6725) (classification)
10-fold cross-validation
Cross Validation Accuracy = 99.3256%

Table 3: Results of the performance tests

4 Conclusions and future perspectives

Besides the good results of the model, much work has to be done in order to create an online FDI system. Scaling the values of the features it is always important to avoid numerical problems during the calculation [6]. In this case it revealed to be crucial. The values of the features resulting from (1) are ten or a hundred times smaller than the others ones. The values with a greater range were dominating the smaller ones, leading to worse results.

The model still need more training in order to detect these and more faults in real time. Other kernels and its parameters must be studied. Techniques like leave-one-out cross-validation should be used in order to understand if more features are needed in the sets.

5 Acknowledgements

This work was financially supported by the Project FCT - ORCHESTRA (PTDC/EMS-CRO/2042/2012) of the Fundação para a Ciência e Tecnologia, Portugal.

References

[1] Y.-W. Chen and C.-J. Lin. Combining svms with various feature selection strategies. In *Feature extraction*, pages 315–324. Springer, 2006.

[2] D. Dehestani, F. Eftekhari, Y. Guo, S. Ling, S. Su, and H. Nguyen. Online support vector machine application for model based fault detection and isolation of hvac system. In *International Journal of Machine Learning and Computing*, volume 1, pages 1–7, 2011.

[3] J. Duarte, L. Rato, P. Shirley, and M. Rijo. Multi-platform controller interface for scada application. In *Proceedings of 18th IFAC World Congress (IFAC2011)*, volume 18, pages 7885–7890. IFAC, Milan, Italy, 2011.

[4] E. Eyjolfsdottir, G. Tilak, and N. Li. Moviegen: A movie recommendation system. Technical report, Computer Science Department, University of California Santa Barbara, 2010.

[5] I. Guyon and A. Elisseeff. An introduction to variable and feature selection. In *The Journal of Machine Learning Research*, volume 3, pages 1157–1182. JMLR. org, 2003.

[6] C.-W. Hsu, C.-C. Chang, and C.-J. Lin. A practical guide to support vector classification. Technical report, Department of Computer Science, National Taiwan University, 2003.

[7] T. Joachims. *Text categorization with support vector machines: Learning with many relevant features*. Springer, 1998.

[8] J. Lemos, F. Machado, N. Nogueira, L. Rato, and M. Rijo. Adaptive and non-adaptive model predictive control of an irrigation channel. In *Networks and Heterogeneous Media*, volume 4, pages 303–324. American Institute of Mathematical Sciences, 2009.

[9] S. Ramaswamy, P. Tamayo, R. Rifkin, S. Mukherjee, C.-H. Yeang, M. Angelo, C. Ladd, M. Reich, E. Latulippe, J. Mesirov, et al. Multiclass cancer diagnosis using tumor gene expression signatures. In *Proceedings of the National Academy of Sciences*, volume 98(26), pages 15149–15154. National Acad Sciences, 2001.

[10] M. Rijo. Local automatic control modes in an experimental canal. In *Irrigation and Drainage System*, volume 17, pages 179–193, 2003.

[11] J. Weston, S. Mukherjee, O. Chapelle, M. Pontil, T. Poggio, and V. Vapnik. Feature selection for svms. In *NIPS*, volume 12, pages 668–674, 2000.

[12] W. Yin, W. Zhang, and X. Sun. A svm-based multiple faults classification scheme design in flight control fdi system. In *Innovative Computing, Information and Control, 2007. ICICIC'07. Second International Conference on*, pages 187–187. IEEE, 2007.

[13] J. Zeng, D. Lu, Y. Zhao, Z. Zhang, W. Qiao, and X. Gong. Wind turbine fault detection and isolation using support vector machine and a residual-based method. In *American Control Conference (ACC), 2013*, pages 3661–3666. IEEE, 2013.

Is it Possible to estimate the Embedding Dimension using FNN?

Jorge Henrique Santos Oliveira
oliveira_jorge@dcc.fc.up.pt

Miguel Coimbra
m_coimbra@dcc.fc.up.pt

Instituto de Telecomunicações

Faculdade de Ciências da Universidade do Porto
Portugal

Abstract

We are motivated by the fact that strange behaviours in time series like PCG, maybe are very well behaved on a hyperspace reconstructed by the delay vectors.
In this hyperspace, the trajectories are restricted in space and self-similarities are maybe present in different scales. In this work , we aim to find dimensionality of this hyperspace using False Nearest Neighbors (FNN) method.

1 PCG Introduction

Cardiac auscultation is the simplest, fastest and cheapest method for heart examination. It provides information about the structural and functional characteristics of the heart using a simple medical device, the stethoscope. Electronic versions of this device are capable of registering and optimizing the quality of auscultation signals, generating what is called a phonocardiogram (PCG) signal. Data collection systems such as DigiScope [1] can obtain PCG signals such as the one depicted in Figure 1. Here we can observe the various components of a heart cycle, including S1 (first heart sound) and S2 (second heart sound). These establish the boundaries of the other two fundamental components of a heart cycle: the systole (period between S1 and S2), and the diastole (period between S2 and S1). S1 and S2 are generated by the opening and closing of the various heart valves and in some auscultations we have the presence of additional sounds such as S3, S4 or murmurs. [2]

Heart sounds segmentation is a fundamental step for extracting useful physiological information from heart sounds. It is typically preceded by a pre-processing step, and followed by feature extraction and classification steps in traditional statistical pattern recognition approaches [3].

This paper is structured as follows: the Embedding Theorem is presented in Section 2, followed by False Nearest Neighbors explanation in Section 3. The Materials and Results conclude the paper in sections 4 and 5.

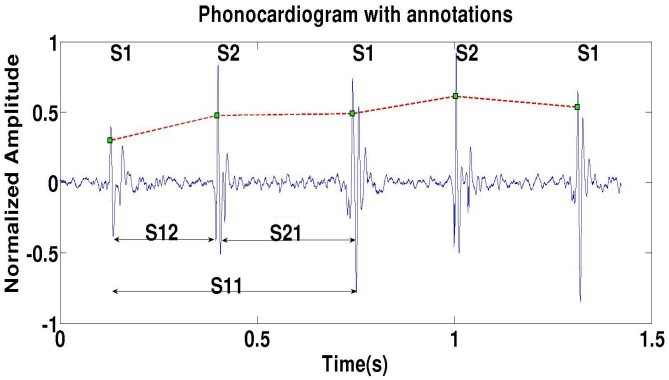


Figure 1: A typical heart sound and its four main components: S1, S2, Systole, Diastole

2 Embedding Theorem

Considering the Embedding Theorem by Takens, one assumes the existence of a embedding dimension m such that the delay vectors $\vec{s}_n$ construct a phase space representing the dynamics of the system [4]. From the geometry of the set in the phase space we want to characterize the dimension of the attractor D which is smaller than the number of degrees of freedom; for fractal dimensionality they are usually called strange attractors [4]. Normally to estimate and embedding dimension the dimensionality of the attractor should be increased until a deterministic behaviour on data appears. Box-counting algorithms can also be used to compute this quantity but turns out to be impractical for systems with $D>2$.

3 False Nearest Neighbors

This method uses the idea of neighbors and two points are considered to be neighbors if their distance is less than a pre-defined threshold. Suppose that the dimension of the delay vector is *m* and it is projected into a lower dimensional m_0 . If the number of false neighbors does not increase substantially from $m \rightarrow m_0$ then it is acceptable to reduce the dimension of the delay vectors [10].

4 Material

This database includes PCGs from pediatric and adult patients with their corresponding S1 and S2 annotated positions. In the pediatric 90 heart sound segments are used corresponding to a total 1415 annotations examples of S1 and S2, within a time range between 1.2 and 14.7 seconds. In the adults 21 heart sound segments are used composing a total 390 annotations, within a time range between 3.4 and 9 seconds.

5 Results

The result is showed in Figure 2 for a sample PCG signal.

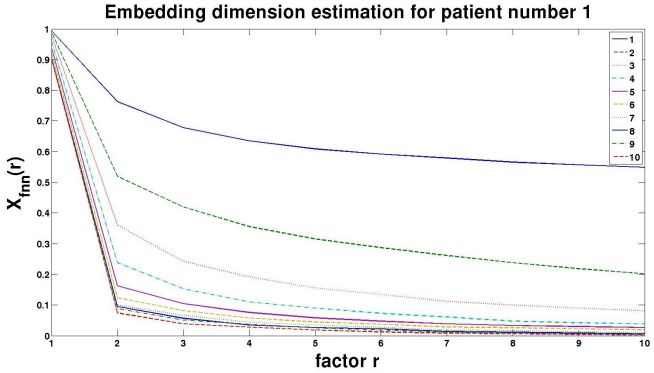


Figure 2: The relative number of false negatives plotted against the minimum distance to be considered as a possible neighbor

Observation of Figure 2 hints that a time window *m* of 25ms (5 samples) is enough. Although beyond the scope of this paper, a deeper inspection of how this varies when additional sounds are present (S3, S4, murmurs) is an interesting topic to be explored in the future.

Acknowledgement

This work was funded by the Fundação para a Ciência e Tecnologia (FCT, Portuguese Foundation for Science and Technology) under the grants Heart Safe PTDC/EEI-PRO/2857/2012 and PEst-OE/EEI/LA0008/2013.

References

[1] D. Pereira, F. Hedayioglu, R. Correia, T. Silva, I. Dutra, F. Almeida, S. S. Mattos, M. Coimbra, "DigiScope – Unobtrusive Collection and Annotating of Auscultations in Real Hospital Environments", in IEEE EMBC, Boston, USA, Sep 2011.

[2] A.Guyton, J.E.Hall, eds., Textbook of Medical Physiology. Elsevier Saunders, 11th ed., Ed Hall, Jun 2006.

[3] H. Kantz, T. Schreiber Nonlinear Time Series Analysis , 2th ed. vol .3, Ed. Cambridge University Press, Jan 2004,cp 3-9.

[4] Vivek.N Roland.P, 08/2005, Simplicity Based Gating of Heart Sounds, *, Circuits and Systems, 2005. 48th Midwest Symposium on,* 1298 - 1301 ,Vol. 2.

[5] P.Bentley, G.Nordehn, M.Coimbra, S.Mannor "The PASCAL Classifying Heart Sounds Challenge 2011 (CHSC2011) Results. " http://www.peterjbentley.com/heartchallenge/

Protein quantification of fluorescence images in heterogeneous cell population

Cláudia Gomes[1] (claudia.c.gomes@tecnico.ulisboa.pt)

Raquel Seruca[2] (rseruca@ipatimup.pt)

João Sanches[1] (jmrs@tecnico.ulisboa.pt)

[1]Institute for Systems and Robotics,
Instituto Superior Técnico, University of Lisbon

[2]Institute of Molecular Pathology and Immunology of the University of Porto

Abstract

For cancer diagnosis and prognosis or functional genetic analysis quantification and mapping of protein expression is mandatory. In several models proteins or vectors are stained with fluorescence images of microscopy are obtained. The analysis of these images is usually qualitative and till now it has been difficult to propose a robust quantification approach to such images. One of the key limitations is due to the irregular nature of shape and size of the cells within the population of interest.

In this paper the authors present an improved version of the previous methodology proposed for quantitative analysis which is able to deal with this population heterogeneity.

The core of the method is an optimization algorithm where the distortions associated with each observed noisy intensity profiles are jointly estimated and inverted in order to compute the unknown ideal expression level that characterize the population.

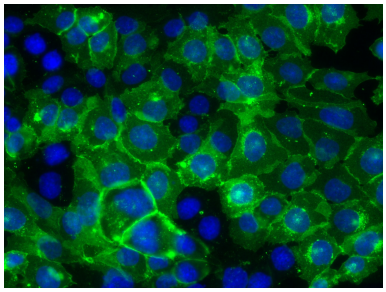


Figure 1 – Immunofluorescence image showing membranous protein expression in a cell population. The protein (E-Cadherin) is labelled with green fluorescence.

1 Introduction

One of the most common methodologies for qualitative analyses of protein expression is *immunofluorescence* (IF) (Fig. 1) which is extensively used in biological and medical research. In all microscopic approaches only molecules tagged with fluorescent dyes are captured [1], allowing the observation of very specific biological processes. Many image processing procedures and algorithms are described for IF images and most of them are related to denoising [2], morphologic analysis [3], cell tracking [4], shape characterization and detection [5] and object counting. In biomedical research, quantitative measures extracted from IF images were rarely explored to identify dysfunctional proteins associated with a particular biological behavior. In this scope, image processing of IF images were only developed for quantification of cell volumes and single cell movement analysis. The common methods for single cell analyses includes *Fluorescence Resonance Energy Transfer* (FRET), *Fluorescence Correlation Spectroscopy* (FCS), *Fluorescence Recovery After Photobleaching* (FRAP) and *Fluorescence Lifetime Imaging Microscopy* (FLIM) which are able to quantify parameters of local dynamic processes within a single cell.

For cell populations, a central difficulty in imaging quantification is the calibration process related to the heterogeneity of parameters occurring in image acquisition, which most of the times is not considered.

In this work we propose a method, applied to a population of cells as shown in Fig. 1, to characterize the level of expression and distribution of tagged fluorescent molecules in the intra and intercellular space. We designed an algorithm that calibrates the morphological variability of a cell population and computes a *one dimensional* (1D) representative profile of protein distribution [7]. This allows the representation and quantification of the typical level of expression and distribution of proteins using IF real images.

From each IF image (Fig. 1) pairs of cells were selected for quantification analysis. This first step intended to exclude all negative cells that may represent technical pitfalls. A specific software application was produced to assist in this task.

The mapping and quantification of the level of protein expression is performed by computing 1D intensity *internuclear* (IN) and *radial* (RD) profiles that represent the typical distribution of the molecule between or within cells. The schematic representation of the IN and RD non-geometric compensated profiles is illustrated in Figure .

The goal of the proposed algorithm is to compensate geometrically the column profiles represented in Figure .b) and d) in order to obtain a new profile map, horizontally invariant, from which the ideal distribution profile of the molecules is computed. The main difference between the algorithm described in this paper and the previous version [7] is related with the representation of the underlying unknown profile. Before, a number of unknown profiles are estimated under the constraint that they should be similar. In this version, a single profile is considered and all the observations are assume to be distorted versions of this single profile. The mathematical formulation in this case is simpler and more elegant.

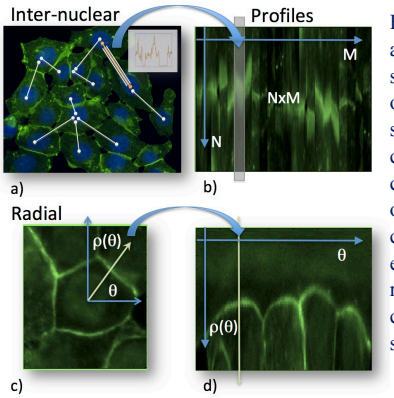


Figure 2 – Extraction scheme of IN and RD profiles. a) Semi-automatic selection procedure of cells and pair of cells (the geometric centre of the selected cells is automatically computed), b) the IN profile map is computed by stacking the profiles obtained from the axis linking the cells of the pair, c) RD profile extraction procedure where a set of radial profiles are extracted from the centre of each individual cell and d) stacked in the RD profile map.

2 Problem Formulation

This section describes the mathematical formulation of the algorithm used to estimate the ideal molecule distribution. The mathematic formulation described in this section aims to create an ideal profile representative of molecule distribution between cells and within each cell. To obtain this ideal representation of the molecule distribution, RD and IN profiles are extracted and a profile map is constructed where each column represented the expression level of the molecule for each angle, in the case of radial profiles, or across the axis linking the cells of each pair, in the case on IN profiles (see Figure).

Let us consider an unknown intensity profile defined as follows

$$f(x) = \sum_{k=0}^{N-1} c_k \phi_k(x) = \Phi^T(x)\, \boldsymbol{c} \tag{1}$$

where $\Phi(x) = [\phi_k(x)]$ is a column vector of N known basis functions and $\boldsymbol{c} = [c_k]$ is a column vector of coefficients to be estimated. Let also consider a set of M radial or inter-nuclear L length observed profiles, $\boldsymbol{p}_j = [\mathbf{x}_j, \mathbf{y}_j]$ with $0 \le j \le M-1$. $\boldsymbol{x}_j = [x_{ij}]$ and $\boldsymbol{y}_j = [y_{ij}]$ are L length column vectors containing the locations and intensities values of the j^{th} observed profile. Intensities y_{ij} are assumed to be noisy and the locations x_{ij} are assumed to be geometrically distorted, $x_{ij} = x_{ij}^* + t_{ij}$ where x_{ij}^* is the true unknown location and t_{ij} is the corresponding compensation shift value.

The ideal profile $\boldsymbol{c}$ and the compensated locations of the observations, stacked in a $L \times M$ matrix containing M observed profiles with L points each, X, are computed by solving the following optimization task

$$[\boldsymbol{c}, X] = \arg_{\boldsymbol{c},X} \, min \; E(Y, X, \boldsymbol{c}) \tag{2}$$

where the energy function is of the form

$$E(Y, X, \boldsymbol{c}) = E_d(Y, X, \boldsymbol{c}) + E_c(\boldsymbol{c}) + E_X(X) \tag{3}$$

with the following data fidelity term is

$$\mathrm{E_d(Y, X, c)} = \sum_{\mathrm{ij}} \left(\mathrm{f(x_{ij})} - \mathrm{y_{ij}}\right)^2. \tag{4}$$

$E_c(\boldsymbol{c})$ and $E_X(X)$ are regularization terms for $\boldsymbol{c}$ and X. The smoothness of $f(x)$ is obtained by difining the prior function for $\boldsymbol{c}$ as $E_c(\boldsymbol{c}) = \left|\left|\Theta \boldsymbol{c}\right|\right|_2^2$ where Θ is a $N \times N$ difference operator. The prior for X is designed in order to impose a limited constraint on the observation displacements in order to prevent its collapse in a single point,

$$E_X(X) = \sum_{ij} (x_{ij} - x_{i-1,j})^2 \tag{5}$$

where x_{ij} and $x_{i-1,j}$ are consecutive neighbours from the same observed profile. $x_{0,j}$ and $x_{L-1,j}$ are not updated during the iterative process and are kept fixed.

The global energy function to be minimized is

$$E(Y,X,\boldsymbol{c}) = Tr\big((\Phi^T\mathbf{c}-\boldsymbol{y})^T(\Phi^T\mathbf{c}-\boldsymbol{y})\big) + \alpha\boldsymbol{c}^T\Psi_N\mathbf{c} + \beta\,\mathrm{Tr}(\mathrm{X}^T\Psi_L\mathrm{X}) \quad (6)$$

where $\Psi_\tau = \Theta_\tau^T\Theta$ is a $\tau\times\tau$ matrix, $\Phi(\mathrm{X}) = \big[\phi_k(x_{ij})\big]$ is a $N\times LM$ matrix, $\boldsymbol{y}$ is a LM column vector containing all the observations of the map image of profiles and Tr denotes the trace operator.

The solution of the optimization problem (2) can be obtained in two steps as follows:

$$\boldsymbol{c}^{t+1} = \arg_c min\, E(Y,\boldsymbol{X}^t,\boldsymbol{c}) \quad (7)$$

$$\boldsymbol{X}^{t+1} = \arg_X min\, E(Y,\boldsymbol{X},\boldsymbol{c}^{t+1}) \quad (8)$$

By computing the stationary point of E with respect to $\boldsymbol{c}$, $\nabla_c E(Y,X,\boldsymbol{c}) = 0$ and with respect to X, $\nabla_X E(Y,X,\boldsymbol{c}) = 0$, after straight manipulation it can be obtained the following expressions for $\boldsymbol{c}$ and X,

$$\boldsymbol{c}^{t+1} = (\Phi(\boldsymbol{X}^t)\Phi(\boldsymbol{X}^t)^T + \alpha\Psi_N)^{-1}\Phi\boldsymbol{y} \quad (9)$$

$$\boldsymbol{X}^{t+1} = \frac{1}{3\beta}Z + \bar{Z} \quad (10)$$

where $z_{ij} = (y_{ij} - f(x_{ij}))\partial f(x_{ij})/\partial x_{ij}$ and $\bar{z}_{ij} = (x_{i-1,j} + x_{i+1,j} + x_{i,j})/3$.

3 Results

The results presented in this section were obtained in Matlab where the proposed algorithm is implemented as well as some pre-processment/normalization steps were performed. The normalization of each profile intensity were computed, for each profile p_j, $p_{j,final} = (P_{j,initial} - \mu)/\sigma$ where μ is the mean of the corresponding profile and σ the standard deviation. The profile lengths, stacked in the non compensated profile maps, were normalized between 0 and 1 for a fixed number of samples using interpolation, as shown in Figures 3.a) and 4.a).

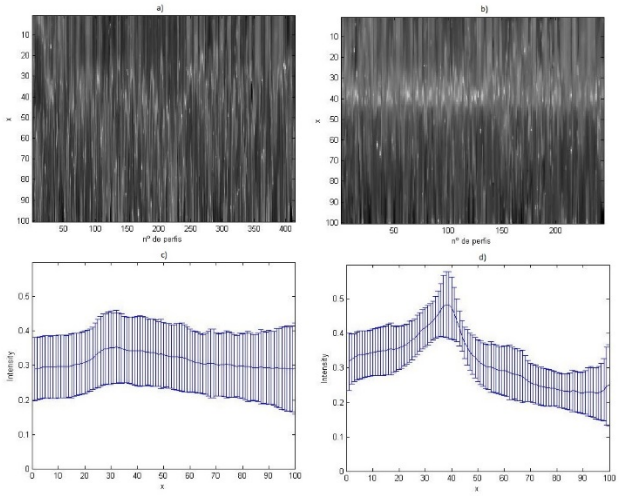


Figure 1: RD profiles map. a) RD profiles map without geometrical compensation, b) RD profiles map with geometrical compensation, c) mean and standard deviation of RD profiles without geometrical compensation and d) mean and standard deviation of IN profiles with geometrical compensation.

For sake of convergence, a mechanism of outliers rejection was used in the iterative process described by equations (7) and (8). The relative difference between each profile and the current estimate of $\boldsymbol{c}^t$, at iteration t, was computed and used as a measure to exclude outliers. When the norm of the difference between the most distant profile and $\boldsymbol{c}^t$ is above a given threshold the profile is considered an outlier and is excluded.

Figure 3 displayed the mean (column average) RD profiles computed from the non geometrically compensated map of profiles (Fig. 3.a)) and from the geometrically compensated map of profiles (Fig.3.b) in Fig.3.c) and Fig.3.d) respectively. In this figures the error bars represent the standard deviation of the estimated average at each location.

Figure 4, represent similar results for the case of IN profiles. The non-compensated and compensated profile maps are shown in Fig.4.a) and b) respectively and the corresponding average profiles are displayed in Fig. 4.c) and d). In the case of IN case it is clear the decrease on standard deviation of the average profile when the compensation algorithm is used.

Figure 5 displays the estimated ideal RD (Fig.5.a)) and IN (Fig.5.b)) profiles obtained with the proposed algorithms. These ideal profiles are similar to the average profiles displayed in Figs. 3.d) and 4.d) but are not equal.

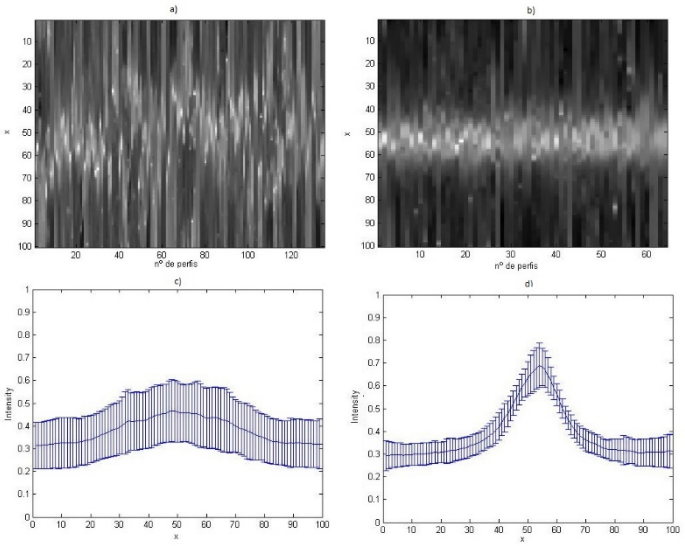


Figure 4: IN profiles map. a) IN profiles map without geometrical compensation, b) IN profiles map with geometrical compensation, c) mean and standard deviation of IN profiles without geometrical compensation and d) Mean and standard deviation of IN profiles with geometrical compensation.

For both RD and IN examples 2000 iterations were performed and the number of basis function used in these simulations is $N = 100$. The prior parameters in the RD case are $\alpha = 150$ and $\beta = 400$ and $\alpha = 200$ and $\beta = 500$ in the case of the IN case.

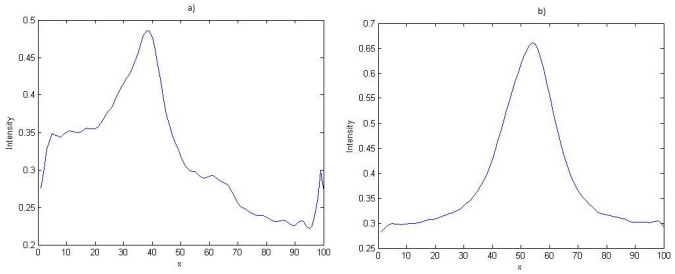


Figure 5: Ideal RD (a) and IN (b) profiles obtained from (7).

4 Conclusions

This paper describes an improvement on the previous work of the authors about estimating an ideal profile representative of the molecule distribution on intra and inter cellular space computed from a population of irregular cells. Tests with real data to illustrate the application of the method were performed in a series of 8 cell lines stably expressing WT protein (E-cadherin) or 7 aberrant proteins (E-cadherin mutant molecules) spread along the protein.

References

[1] I. D. Odell and D. Cook, "Immunofluorescence techniques," Journal of Investigative Dermatology, vol. 133, no. 1, p. e4, 2013.

[2] I. C. Rodrigues and J. M. R. Sanches, "Convex total variation denoising of poisson fluorescence confocal images with anisotropic filtering," Image Processing, IEEE Transactions on, vol. 20, no. 1, pp. 146–160, 2011.

[3] E. Meijering, "Cell segmentation: 50 years down the road [life sci- ences]," Signal Processing Magazine, IEEE, vol. 29, no. 5, pp. 140–145, 2012.

[4] A. Mosig, S. Jger, C. Wang, S. K. Nath, I. Ersoy, K. Palaniappan, and S.-S. Chen, "Tracking cells in life cell imaging videos using topological alignments." Algorithms for Molecular Biology, vol. 4, no. 10, July 2009.

[5] C. Ducroz, J.-C. Olivo-Marin, and A. Dufour, "Characterization of cell shape and deformation in 3d using spherical harmonics." in ISBI. IEEE, 2012, pp. 848–851.

[6] D. S. Baskin, M. A. Widmayer, and M. A. Sharpe, "Quantification and calibration of images in fluorescence microscopy," Analytical Biochem- istry, vol. 404, no. 2, pp. 118 – 126, 2010.

[7] S. Esménio, J. Figueiredo, R. Seruca, and J. M. Sanches, "E-cadherin radial distribution characterization for mutation detection purposes," in IbPRIA, 2013, pp. 173–180.

Improving Classification Accuracy of Deep Neural Networks by Transferring Features from a Different Distribution

Chetak Kandaswamy[1]
chetak.kand@gmail.com
Luís M. Silva[1]
lmas@ua.pt
Jaime S. Cardoso[2]
jaime.cardoso@inescporto.pt

[1] Instituto de Engenharia Biomédica (INEB),
Rua do Campo Alegre, 823
Porto, Portugal

[2] INESC TEC and Faculdade de Engenharia,
Universidade do Porto, Rua Dr. Roberto Frias, 378
Porto, Portugal

Abstract

Deep Neural Networks (DNN) are feed-forward, artificial neural networks that allow learning of multiple levels of abstraction that help to make sense of data such as images, sound, and text. We study the performance of DNNs using transfer learning approaches. Transfer learning is a process where a network trained on a source problem is reused to solve a new target problem by applying minor modifications to the network. Generally, transfer learning is done when the distributions between the source and target are similar and the task is equal. In this paper, we hypothesis that, if the distance between the distributions is different, and the tasks are different transfer learning will help to improve classification performance and speed up the process.

For this purpose we propose unsupervised feature transference using Stacked Denoising Autoencoder (SDA). In unsupervised feature transference approach we explore: 1) transfer learning between completely different tasks drawn from different distributions and, 2) transfer learning between equal tasks drawn from different distributions. We achieved significant improvement on average error rate and on average computation time using SDA on two types of transfer learning approaches to test our hypothesis.

1 Introduction

The study of transfer learning was inspired by the ability of humans to reuse prior experience under different environments. Naturally, the transfer learning paradigm implies reusing learning machines previously trained for a given source problem S in order to solve, with minor modifications, a different target problem T. An ideal transfer learning method should improve the reused classifier over the one trained from scratch.

Traditionally, transfer learning is applied between problems with same classification task and drawn from similar distributions. In this paper we primarily explore transfer learning between *completely different tasks* drawn from *different distributions* [1], i.e., a classifier learns to perform a task on training instances drawn from the source problem and then reused to perform a different task on a target problem instances drawn from a different distribution. Secondly, we explore transfer learning between problems with same classification task but drawn from different distributions.

We analyze Deep Transfer Learning (DTL) (transfer learning with deep architectures) using an unsupervised feature transference (USDA) approach [1]. We use state-of-the-art deep learning methods like SDA (see [2],) that learn high-level features from large datasets. In order to distinguish how different the target distribution is from the source distribution, we use Jensen-Shannon divergence (JSD) as a metric to measure the degree of heterogeneity between distributions.

2 Problem Formulation: Transfer Learning

Given an input space X and a set of labels Y, a classifier is any function $g(\mathbf{x}) : X \rightarrow Y$ that maps instances $\mathbf{x} \in X$ to labels. Essentially, Y is a coding set for the labels using some one-to-one mapping (e.g., $\Omega = \{\text{“}equilateral\text{”}, \text{“}circle\text{”}, \text{“}square\text{”}\} \rightarrow Y = \{0,1,2\}$ with number of labels $c = 3$). We assume that n_{ds} instances are drawn by an i.i.d. sampling process from the input space X with a certain probability distribution $P(X)$, thus giving a design data set $X_{ds} = \{\mathbf{x}_1, \ldots, \mathbf{x}_{n_{ds}}\}$ which is accompanied by a set of label codes $Y_{ds} = \{y_1, \ldots, y_{n_{ds}}\}$ for each instance. The classifier performance is measured using error rate ε and computation time on a test set $X_{ts} = \{\mathbf{x}_1, \ldots, \mathbf{x}_{n_{ts}}\}$ with n_{ts} unlabeled instances drawn from the same distribution $P(X)$.

Traditionally, the goal of transfer learning is to transfer the learning (knowledge) from a source problem input space X_S to one or more problems or distributions to efficiently develop an effective hypothesis for a new task, problem or distribution. In this framework of transfer learning, the source and target problems may come from equal or different distributions. In supervised learning problems, the source Y_S and target Y_T labels may be equal or different. We focus mainly on two cases where the distributions are different:

1. **(TL: case 1)** The distributions are different $P_S(X) \neq P_T(X)$ and the labels are different $Y_S \neq Y_T$.

2. **(TL: case 2)** The distributions are different $P_S(X) \neq P_T(X)$ and the labels are equal $Y_S = Y_T$.

Under such hypothesis, our goal is to obtain an accurate classification for target-domain instances by exploiting labeled training instances from the source-domain.

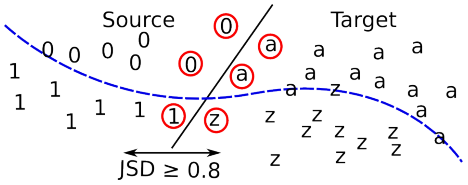


Figure 1: (TL: Case 1) Classify letters reusing digits

Comparing distributions: We use JSD as a measure to compute the difference between the distributions of two datasets. Given two probability functions $P_S(X)$ and $P_T(X)$, when $D_{JS}(S||T) = 0$, we can consider that two distributions are identical and when $D_{JS}(S||T) < 0.5$, the distributions are similar and when $D_{JS}(S||T) \geq 0.5$, the distributions are different.

2.1 Datasets

We use seven datasets of color and gray scale images, as described in Table 1. These seven datasets are either distinct in number of labels or distributions.

Latin and Arabic datasets are representative names for the well-known MNIST[1] and MADbase[2] datasets of handwritten Latin and Arabic digits, respectively. The original Chars74k[3] dataset of typed characters was resized to 28×28 pixels and then broken into three smaller datasets: Digits dataset contains digits from 0-to-9, the Lowercase dataset contains lowercase letters from a-to-z and finally, the Uppercase dataset contains uppercase letters from A-to-Z.

Shape1 and Shape2 are from Baby AI shape dataset[4] which was used for shape recognition tasks. Shape1 consists of images of canonical shapes (equilateral triangles, circles and squares) while Shape2 consists of images of non-canonical shapes (ellipsis, rectangles and triangles). Both Shape1 and Shape2 were generated with 28×28 pixels, with variation of colors: 0 to 7, position: left extreme to right extreme, rotation: 0 to 360^o.

Network Architecture: The SDA network had three hidden layers and one output layer with [576, 400, 256, c] neurons amounting to 784,384 connections. The induced random corruption levels for each of the three

[1] http://yann.lecun.com/exdb/mnist/
[2] http://datacenter.aucegypt.edu/shazeem/
[3] We acknowledge Microsoft Research India for Chars74k dataset.
[4] http://www.iro.umontreal.ca/~lisa/twiki/bin/view.cgi/Public/BabyAIShapesDatasets

Table 1: Dataset characteristics with number of instances.

Data set		Labels			Instances		
Distribution		Ω		c	Train	Validation	Test
Latin	P_L	0-to-9	Ω_{09}	10	50,000	10,000	10,000
Arabic	P_A	•-to-9	$\Omega_{\bullet 9}$	10	50,000	10,000	10,000
Latin-2	P_{L2}	0-to-9	Ω_{09}	10	13,208	6,604	10,000
Digits	P_D	0-to-9	Ω_{09}	10	5,080	2,540	2,540
Lowercase	P_{LC}	a-to-z	Ω_{az}	26	13,208	6,604	6,604
Uppercase	P_{UC}	A-to-Z	Ω_{AZ}	26	13,208	6,604	6,604
Shape1	P_{Sh1}	'eqt','cir','sqr'	Ω_{sh1}	3	10,000	5,000	5,000
Shape2	P_{Sh2}	'tri','ell','rec'	Ω_{sh2}	3	10,000	5,000	5,000

hidden layers inputs are [10%, 20%, 30%] respectively. We used pre-training and fine-tuning learning rates of 0.001 and 0.1, respectively. The stopping criteria for pre-training was fixed to 40 epochs; stopping criteria for fine-tuning was set to a maximum of 1000 epochs with the validation dataset. Each of these experiments is repeated 10 times and performed student t-test with confidence interval of 0.05 to give some statistical significance.

We used Theano, a GPU compatible machine learning library to perform all our experiments on a i7-377 (3.50GHz), 16GB RAM and GTX 770 GPU processor.

3 Proposed approach and results

We propose a feature transference approach which enables deep neural networks to transfer hidden layers features (weights) of a classifier trained in unsupervised way [1]. For that purpose, SDA's are trained on a source problem and its features transferred to help in solving a target problem. We represent this transference by $\mathbf{w}_S \Rightarrow \mathbf{w}_T$. In this paper we explore feature transference in SDA at the pre-training stage, denoted $U(\mathbf{w})$, a process we call unsupervised feature transference (USDA) [1].

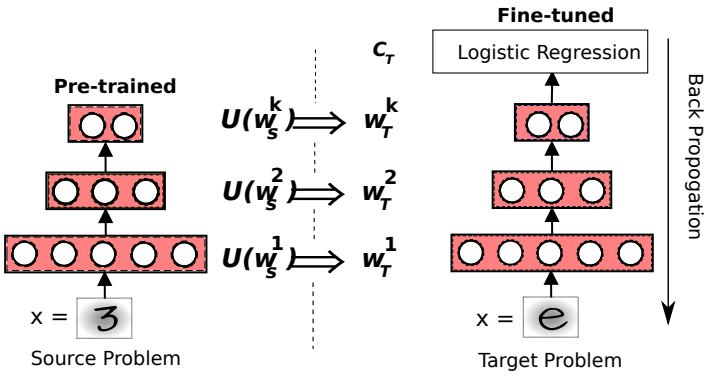


Figure 2: Unsupervised feature transference using SDA

In the USDA approach we transfer the unsupervised features of the SDA model from the source to the target problem, i.e., $U(\mathbf{w}_S) \Rightarrow \mathbf{w}_T$ as depicted in Fig.2. Once the features are transferred to the target problem, we add a logistic regression layer on top of the transferred machine. We fine-tune this entire classifier as a multi-layer perceptron using back-propagation. We compare the performance of USDA approach (TL) to the one obtained from a classifier built from scratch for the target problem, the baseline approach (BL).

3.1 TL: case 1

Classify letters reusing digits: We classify images of typed letters by reusing unsupervised features of a machine trained with handwritten digits from 0-to-9.

The performance is listed in Table 2. The average error rate of recognizing uppercase letters, 4.31±0.16% by reusing a machine pre-trained with Latin digits is significantly *lower* than baseline, 5.01±0.27%. Similar results are obtained from recognizing the lowercase letters. In both cases the significance level allows rejecting the null hypothesis of equal error rates.

It is interesting to note the difference between, reusing either Arabic or Latin dataset. The Latin dataset with smaller JSD value leads to better classification performance than the Arabic dataset. It would be interesting to study, does smaller JSD value leads to better TL even when the tasks are different.

Table 2: Average classification test error using SDA (%) ($\overline{\varepsilon}$) obtained for a target problem using UFT approach for different combinations of: target data distribution (P_T); target label set (Ω_T); source distribution (P_S); source label set (Ω_S); The difference between distributions is given by Jensen-Shannon divergence (JSD).

Label space		P_S	Ω_S	P_T	Ω_T	$\overline{\varepsilon}$		JSD
$Y_S \neq Y_T$	BL			P_{UC}	Ω_{AZ}	5.01±0.27		
	TL	P_L	Ω_{09}	P_{UC}	Ω_{AZ}	**4.31±0.16**	⇑	0.79
	TL	P_A	$\Omega_{\bullet 9}$	P_{UC}	Ω_{AZ}	4.41±0.22	⇑	0.99
	BL			P_{LC}	Ω_{az}	4.95±0.16		
	TL	P_L	Ω_{09}	P_{LC}	Ω_{az}	**4.37±0.13**	⇑	0.80
	TL	P_A	$\Omega_{\bullet 9}$	P_{LC}	Ω_{az}	4.43±0.11	⇑	0.99
$Y_S = Y_T$	BL			P_D	Ω_{09}	1.88±0.14		
	TL	P_L	Ω_{09}	P_D	Ω_{09}	1.78±0.21	○	0.88
	TL	P_A	$\Omega_{\bullet 9}$	P_D	Ω_{09}	**1.75±0.21**	○	0.99
	BL			P_{Sh1}	Ω_{sh1}	**7.88±0.93**		
	TL	P_{Sh2}	Ω_{sh2}	P_{Sh1}	Ω_{sh1}	7.96±0.93	○	0.99

⇑, ⇓, ○ statistically significant improvement or degradation or no change than baseline. The best $\overline{\varepsilon}$ obtained for a target dataset is marked in bold.

3.2 TL: case 2

Classify typed digits reusing handwritten digits: We classify images of typed digits by reusing unsupervised features of a machine trained with handwritten digits from 0-to-9. We observe slight performance improvement in the average error rate of recognizing typed digits.

Classify canonical shapes reusing geometrical shapes: We classify images of geometrical shapes (Shape2 dataset) reusing unsupervised features of a machine trained with canonical shapes (Shape1 dataset). We observe slight performance degradation in the average error rate of recognizing canonical shapes as listed in Table 2.

4 Conclusions

We analyzed the performance of USDA approach on two types of transfer learning approaches. TL case 1: when the source and target distributions and their labels are different. For example, the Jensen-Shannon distance for Arabic dataset with typed lowercase dataset is 0.99 and for Latin dataset with typed lowercase dataset is equal to 0.80, in both cases we observe significant improvement using transfer learning approach. It supports our hypothesis that, reusing the source dataset for the target dataset, even when the distributions and tasks are different, USDA approach helps improve classification performance of the target problem (and also around 50% reduction in computation time, see [1]) . Similar results were observed in case of uppercase dataset, this strengthens our previous argument.

On the other hand, TL case 2: when the source and target distributions are different yet the tasks are same, we observe slight improvement using transfer learning approach, i.e., reusing handwritten digits to classify typed digits.

Acknowledgment

The authors would like to thank Neural Network Interests Group of INEB for their valuable and constructive suggestions during the development of this research work. This work was financed by FEDER funds through the Programa Operacional Factores de Competitividade – COMPETE and by Portuguese funds through FCT – Fundação para a Ciência e a Tecnologia in the framework of the project PTDC/EIA-EIA/119004/2010.

References

[1] Chetak Kandaswamy, Luís M Silva, Luís A Alexandre, Ricardo Sousa, Jorge M Santos, and Joaquim Marques de Sá. Improving transfer learning accuracy by reusing stacked denoising autoencoders. *SMC conference, IEEE*, 2014.

[2] Pascal Vincent et al. Stacked denoising autoencoders: Learning useful representations in a deep network with a local denoising criterion. *The Journal of Machine Learning Research*, 11:3371–3408, 2010.

Transfer Learning: Current Status, Trends and Challenges

Ricardo Sousa[1]
rsousa@rsousa.org
Luis M. Silva[2]
lmas@up.pt
Luis A. Alexandre[3]
lfbaa@di.ubi.pt
Jorge Santos[4]
jms@isep.ipp.pt
Joaquim Marques de Sá[1]

[1] INEB, Porto, Portugal
[2] INEB, Porto, Dep. de Matemática at Universidade de Aveiro, Aveiro, Portugal
[3] IT, Universidade da Beira Interior, Covilhã, Portugal
[4] INEB, Porto, Instituto Superior Engenharia do Porto, Porto, Portugal

Abstract

Transfer Learning (TL) has gained significant interest in the Machine Learning (ML) community. Aiming to overcome standard ML learning models assumption that data distributions are the same independently of its origins, a wide variety of works have emerged to tackle this issue. Nowadays, with data available on large amounts and from different applications, and where labeling is a cumbersome task, it is counter productive to have intelligent systems for each specific real-world problem: DVD and home appliances rating, Amazon and medical reports analysis or histology imaging. In this paper we will review current work on TL, open issues and challenges that have not been addressed yet. We conclude this work with a set of remarks and guidelines for future TL methodologies.

1 Introduction

One common assumption in statistical learning theory when devising learning algorithms is that data from different problems are drawn from the same underlying distribution. However, this assumption fails in many real world problems. For the particular case of product reviews, models that perform recommendations of goods on the Amazon website cannot be straightforward applied in the IMDB website. In the same way, data (e.g., text statistics) present in the Wikipedia cannot correspond to the same information as data from the Reuters website. Moreover, despite the fact that textual data is presented in English – and the established grammatical rules are applied – a review that provides a positive feedback for a home appliance device cannot be derived in the same way as a positive review for a horror film. The interest on Transfer Learning (TL) is multifold. First and foremost, it alleviates the need of data labeling which is an expensive and cumbersome job; second, it often produces algorithms with good generalization capability for different learning problems. Finally, it has been claimed that TL provides learning models with good generalization performances in different problems with far less computational effort [4, 16, 17]. In a nutshell, it is the ability of *reusing* what was learned on one problem (also coined as source problem) onto another (target problem). Reusability will be the cornerstone of this manuscript. So, what is TL?

Definition 1 (TL): *TL is a Machine Learning (ML) research field whose goal is the development of algorithms capable of transferring the learning model obtained in a source problem to a target problem without the necessity of building a new model from scratch.*

For more than three decades there has been a significant amount of work on TL. Strangely enough, however, recent and classical works on TL have not been considered or thoroughly analyzed. Moreover, most of the times the concept of TL has been mixed with active, online and even sequential learning [32]; or, concepts from statistical classical learning theory have been used to define all possible TL scenarios. TL in fact shares ideas from areas such as *dataset shift* where the distribution of the data can change over time (and to which sequential algorithms may be applied). However, dataset shift is part of TL mostly because of the assumptions on the data distribution [25, Chapter 1]. Overall, some of these discrepancies have been hindering the evolvement of TL. In this work we will analyze the current developments of TL and open issues identified until now.

2 Transfer Learning: Current Status

To realize how much work has been done in TL, we have to go back some decades. In fact, this subject has been around since the 80's with considerable advancements since then (see for instance [4, 10, 14, 21, 23, 30, 31] and references therein). Probably, the foremost work that envisioned the concept of TL was Tom Mitchel [21] where the idea of bias learning was first presented. A first attempt to extend these ideas was soon performed in [24] where Neural Networks (NNs) where first used for TL. Their pragmatic approach consisted on training a NN on a source problem to be then retrained on a target problem. In a simple way, layer weights obtained using the source problem data were then reused and retrained to solve the target problem. At the time based on Decision Trees, Pratt adopted entropy measures to assess the quality of the hyperplanes. Soon after, [15] derived a framework to use (abstract) internal representations generated by NNs for TL problems.

After these pivotal works, a significant number of implementations and derivations of TL started to appear. In [29] a new learning paradigm was proposed for TL where one would inclemently learn concept after concept. Sebastien Thrun envisioned this approach to how humans learn: by stacking knowledge upon another (as building blocks) resulting in an extreme nested system of learning functions. At that time, a particular case of [29], coined as Multi-Task Learning (MTL), was presented [8, 9] along with their theoretical formulations [2]. In a nutshell, MTL solves all target problems all at once. However, this approach does not hold for our definition of TL (see **Definition 1**) since it learns a common representation for all of our data. These approaches assume that there is a significant amount of information overlapping all concepts that need to be learned, which sometimes is not the case.

In 2000 a specific formulation was introduced in [27] by Shimodaira. Although initially not contextualized in the domain of TL, its theoretical conclusions on how to learn a model on a target problem based on a source problem had a significant impact and its implications were only later realized. [27] described a weighted least squares based on the prior knowledge of the densities of source and target problems. At the time, Shimodaira only addressed the issue of data probabilities being different leading to what he had termed *covariate shift*. After that, [3, 10, 13] presented different algorithms to address the limitations of [27] such as the estimation of data probabilities leading to the rise of the *domain adaptation*[1]. In [28] an extension of Shimodaira's work was presented so that it could cope with the leave-one-out risk. The implications of this specific trend of TL was on the resolution of many Natural Processing Language (NPL) problems [3, 5] and genome sequence analysis [26]. Recently, an overview on TL was presented in [22] with a vast, but horizontal, analysis of the most recent works that tackle classification, regression and unsupervised learning for TL. [18] provided fundamental mathematical reasonings for TL by devising: 1) generalization bounds for max-margin algorithm such as SVMs and 2) its theoretical limits for error variability based on the leave-one-out risk [6]. [18], to the best of our knowledge, was one of the first to identify a gap in the literature of the theoretical limitations of algorithms on TL.

NNs are the majority of the chosen algorithms to perform TL. With the recently re-interest on NNs and the availability of more computational power along with new and faster algorithms, NN with deep architectures started to emerge to tackle TL. In [14] a framework for covariate shift with deep networks was presented. [1, 16, 17] widened the research line of [24] by addressing the following questions: How can we tailor deep neural networks for TL? How TL performs with reusing layers and with

[1] Although both terms are widely used in the literature, the underlying principles are strictly the same. For this reason we will opt to use instead covariate shift to avoid confusion on the terminology.

different types of data? As described so far, in spite of the immense different TL interpretations and definitions, concerns started to appear on how to unify this area of research on ML. Sharing these views, in [23] an unifying framework for many of the existing TL methods is presented. Here, concepts for covariate shift and, in more general, TL, are jointly defined.

3 Transfer Learning: Trends and Challenges

We will now focus on other pressing issues that ought be addressed in the near time soon. One important aspect is how to measure knowledge gains when doing TL. We will also briefly address the impact that differences among datasets (source and target problem) may have in the learning rates. Other open issues are concerned with the increasing availability of data. Knowing that is humanly impossible to analyze this amount of data we found that there are few learning models that address this; finally, we also identified a gap in public competitions to benchmark available methods that are presented in the literature.

Unification of TL: One main issue that has been hindering the advance of TL is the vast amount of formulations on TL. For instance, we have shown the work of [24, 29] that promote the idea of never-ending learning, the covariate shift of Shimodaira [27] and domain adaptation of [3, 4]. They share concepts from Shimodaira's covariate shift, but variants on the terminology are employed which inevitably leads to confusion. In fact, a first tentative for an unification on TL was proposed in [23].

Measuring Knowledge Gains: Bengio *et al*. in [14] analyzed until a certain extent how to quantify TL gains. Although overcoming some interpretation issues regarding performance results that can occur when dealing with multiple source domains, it is unknown how these measures behave in other TL methods besides covariate shift, particularly in situations where class sets are different between problems. Simpler approaches like mean squared error (MSE) or statistical inspired coefficients can provide further information such as class agreement. Moreover, measures as the ones employed by Bengio in [14] can lead to non definite results if one obtains a perfect baseline model (see [14]).

Dissimilar Datasets: How difficult is to do TL? It is important to state that according to [12] the quality of the results is related with Kullback-Leibler divergence measured on the datasets with different source/ target problem pair. In a straightforward reading it may seem that for different problems it may be infeasible to perform TL; Or, that TL models need to be more robust for heterogeneous problems; Or, that data features are not representative. Based on our review, it was not possible to identify works that try to make this analysis or at least to perform an attempt on that. Although these intuitive ideas have empirically present a relation between domain divergence and TL algorithms performance, a theoretical reason for these behaviors is still unknown [7, 17].

A new trend: Big Data With the emergence of evermore data it is infeasible for a human to analyze it in its life time. There has been a special interest in big data, particularly in the biology research fields [19]. However, the development of learning models for each specific problem may be a greedy and slow, but unnecessary, endeavor. Besides the fast-paced research for the development of new learning models to deal with big data, TL still seems to be very far behind. Due to the large size of these datasets, it is also a cumbersome and counter-productive task to pre-process it. Therefore, highly heterogeneous, with under-represented classes and contaminated with noise data may be difficult to process by an automatic learning algorithm. To overcome this, authors choose to use reduced, cleaner, versions of these datasets. See for instance [14]. We see that big data trend only strengthens the motivations and objectives for TL. One of the first works that tackle this was presented in 2014 in [11].

Public Competitions for Learning Algorithms Benchmark Contrary to many other research fields such as fingerprints[2] or object categorization[3], there has been a void in open challenges in TL where one can benchmark our method with others from the literature. Even though there have been workshops and special sessions in international top tier conferences such as 2013 and 2011 edition of NIPS[4], 2011 edition of ICML[5], or, 2009 edition of ECML PKDD[6] where several fundamental questions on TL were addressed and pushed forward the research in this field, to the best of our knowledge, the only competitions on TL were at ICML [20] and NIPS[7].

4 Conclusions

At this point it should be clear the interest on Transfer Learning (TL). We have described the limitations of standard machine learning (ML) approaches and how TL can aid in this matter. We have identified key, breakthrough, works on this subject but there are still aspects that need to be addressed. Most of the fundamental issues on TL such as the assumption on data distribution, generalization bounds or loss functions have started to be explored. In fact, some of these issues were raised in [20], but few authors have researched this. Despite of the amount of works that have been presented we have only seen a glimpse of what can be done on TL. Public competitions had always been a way to present breakthrough results and to assess in a homogeneous manner the state-of-the-art on various domains of research. But, only two competitions were conducted so far. In the future, it should be clearer in what TL consists by unifying its principles and concepts, by providing means to understand the impact of different pairs of source/target problems on the performance results, and finally, ways to measure it.

References

[1] Telmo Amaral, Luis M. Silva, Luis M. Alexandre, Chetak Kandaswamy, Joaquim Marques de Sá, and Jorge Santos. Transfer learning using rotated image data to improve deep neural network performance. In *ICIAR*, 2014.

[2] Jonathan Baxter. A model of inductive bias learning. *J. Artif. Intell. Res.(JAIR)*, 12:149–198, 2000.

[3] Shai Ben-David, John Blitzer, Koby Crammer, Fernando Pereira, et al. Analysis of Representations for Domain Adaptation. *Advances in neural information processing systems*, 19:137, 2007.

[4] Shai Ben-David, John Blitzer, Koby Crammer, Alex Kulesza, Fernando Pereira, and Jennifer Wortman Vaughan. A theory of learning from different domains. *Machine Learning*, 79(1-2):151–175, October 2009. ISSN 0885-6125. doi: 10.1007/s10994-009-5152-4.

[5] John Blitzer, Mark Dredze, and Fernando Pereira. Biographies, bollywood, boom-boxes and blenders: Domain adaptation for sentiment classification. In *ACL*, volume 7, pages 440–447, 2007.

[6] Olivier Bousquet and André Elisseeff. Stability and generalization. *The Journal of Machine Learning Research*, 2:499–526, 2002.

[7] Lorenzo Bruzzone and Mattia Marconcini. Domain adaptation problems: a DASVM classification technique and a circular validation strategy. *IEEE transactions on pattern analysis and machine intelligence*, 32(5):770–87, May 2010. ISSN 1939-3539. doi: 10.1109/TPAMI.2009.57.

[8] Caruana. Multitask Learning. *Machine Learning*, 75:41–75, 1997.

[9] Caruana. *Multitask Learning*. PhD thesis, 1997.

[10] Olivier Chapelle, Bernhard Schölkopf, Alexander Zien, et al. *Semi-supervised learning*, volume 2. MIT press Cambridge, 2006.

[11] Zhiyuan Chen and Bing Liu. Topic modeling using topics from many domains, lifelong learning and big data. In *Proceedings of the 31st International Conference on Machine Learning (ICML-14)*, pages 703–711, 2014.

[12] Wenyuan Dai, Qiang Yang, Gui-Rong Xue, and Yong Yu. Boosting for Transfer Learning. In *Proceedings of the 24th International Conference on Machine Learning*, ICML '07, pages 193–200, New York, NY, USA, 2007. ACM. doi: 10.1145/1273496.1273521.

[13] Hal Daumé III and Daniel Marcu. Domain adaptation for statistical classifiers. *J. Artif. Intell. Res.(JAIR)*, 26: 101–126, 2006.

[14] Xavier Glorot, Antoine Bordes, and Yoshua Bengio. Domain adaptation for large-scale sentiment classification: A deep learning approach. In *Proceedings of the 28th International Conference on Machine Learning (ICML-11)*, pages 513–520, 2011.

[15] Nathan Intrator. Making a Low-dimensional Representation Suitable for Diverse Tasks. *Connection Science*, 8 (2):205–224, 1996. doi: 10.1080/095400996116884.

[16] Chetak Kandaswamy, Luis M. Silva, L. M. Alexandre, Jorge Santos, and JP Marques de Sá. Improving deep neural network performance by reusing features trained with transductive transference. In *ICANN*, 2014.

[17] Chetak Kandaswamy, Luis M. Silva, L. M. Alexandre, Ricardo Sousa, Jorge Santos, and JP Marques de Sá. Improving transfer learning accuracy by reusing stacked denoising autoencoders. In *Proceedings of the IEEE SMC Conference*, 2014.

[18] Ilja Kuzborskij and Francesco Orabona. Stability and hypothesis transfer learning. In *Proceedings of The 30th International Conference on Machine Learning*, pages 942–950, 2013.

[19] Vivien Marx. Biology: The big challenges of big data. *Nature*, 498(7453):255–260, 2013.

[20] Grégoire Mesnil, Yann Dauphin, Xavier Glorot, Salah Rifai, Yoshua Bengio, Ian J Goodfellow, Erick Lavoie, Xavier Muller, Guillaume Desjardins, David Warde-Farley, et al. Unsupervised and transfer learning challenge: a deep learning approach. In *ICML Unsupervised and Transfer Learning*, pages 97–110, 2012.

[21] Tom M Mitchell. The need for biases in learning generalizations. Department of Computer Science, Laboratory for Computer Science Research, Rutgers Univ., 1980.

[22] Sinno Jialin Pan and Qiang Yang. A Survey on Transfer Learning. *IEEE Transactions on Knowledge and Data Engineering*, 22(10):1345–1359, October 2010. ISSN 1041-4347. doi: 10.1109/TKDE.2009.191.

[23] Novi Patricia and Barbara Caputo. Learning to learn, from transfer learning to domain adaptation: A unifying perspective. In *Proceedings of the Computer Vision and Pattern Recognition*, 2014.

[24] Lorien Y Pratt, LY Pratt, SJ Hanson, CL Giles, and JD Cowan. Discriminability-Based Transfer between Neural Networks. In S J Hanson, J D Cowan, and C L Giles, editors, *Advances in Neural Information Processing Systems 5*, pages 204–211. 1992.

[25] Joaquin Quionero-Candela, Masashi Sugiyama, Anton Schwaighofer, and Neil D Lawrence. *Dataset shift in machine learning*. The MIT Press, 2009.

[26] Gabriele Schweikert, Gunnar Rätsch, Christian Widmer, and Bernhard Schölkopf. An empirical analysis of domain adaptation algorithms for genomic sequence analysis. In *Advances in Neural Information Processing Systems*, pages 1433–1440, 2009.

[27] Hidetoshi Shimodaira. Improving predictive inference under covariate shift by weighting the log-likelihood function. *Journal of statistical planning and inference*, 90(2):227–244, 2000.

[28] Masashi Sugiyama, Matthias Krauledat, and Klaus-Robert Müller. Covariate shift adaptation by importance weighted cross validation. *The Journal of Machine Learning Research*, 8:985–1005, 2007.

[29] S. Thrun. Is Learning the *n*-th Thing Any Easier Than Learning the First? In D. Touretzky and M Mozer, editors, *Advances in Neural Information Processing Systems (NIPS) 8*, pages 640–646, Cambridge, MA, 1996. MIT Press.

[30] T Tommasi, F Orabona, and B Caputo. Learning Categories from Few Examples with Multi Model Knowledge Transfer. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, PP(99):1, 2013. ISSN 0162-8828. doi: 10.1109/TPAMI.2013.197.

[31] V N Vapnik. An overview of statistical learning theory. *IEEE transactions on neural networks / a publication of the IEEE Neural Networks Council*, 10(5):988–99, January 1999. ISSN 1045-9227. doi: 10.1109/72.788640.

[32] Peilin Zhao and Steven C Hoi. OTL: A framework of online transfer learning. In *Proceedings of the 27th International Conference on Machine Learning (ICML-2010)*, pages 1231–1238, 2010.

[2] FVC ongoing contest: https://biolab.csr.unibo.it/FVCOnGoing/

[3] VOC Challenge: http://pascallin.ecs.soton.ac.uk/challenges/VOC/

[4] NIPS TL: http://nips.cc/Conferences/2013/Program/event.php?ID=3721, and https://sites.google.com/site/nips2011domainadap/

[5] ICML: http://clopinet.com/isabelle/Projects/ICML2011/

[6] ECML: http://www.ecmlpkdd2009.net/program/tutorials/transfer-learning-for-reinforcement-learning-domains/

[7] NIPS TL competition: https://sites.google.com/site/nips2011workshop/transfer-learning-challenge

CorSiL: A Novel Dataset for Portuguese Sign Language and Expressiveness Recognition

Pedro M. Ferreira[1]
pmmf@inesctec.pt
Inês V. Rodrigues[1]
ines.vigario@gmail.com
Ana Rio[4]
anaespinheira@hotmail.com
Ricardo Sousa[2]
rsousa@rsousa.org
Eduardo M. Pereira [1,3]
ejmp@inesctec.pt
A. Rebelo[1]
arebelo@inesctec.pt

[1] INESC TEC
Rua Doutor Roberto Frias, 378, 4200-465
Porto, Portugal

[2] INEB - Instituto de Engenharia Biomédica
Rua do Campo Alegre, 823, 4150-180
Porto, Portugal

[3] Faculdade de Engenharia da Universidade do Porto
Rua Dr.Roberto Frias, s/n, 4200-465
Porto, Portugal

[4] Agrupamento de Escolas Eugénio de Andrade
Rua Augusto Lessa, 4200-098
Porto, Portugal

Abstract

One of the main challenges in the development of any automatic recognition system, specially in the sign language field, is the availability of suitable ground-truth data. In this paper, a novel video-based database, called CorSiL, is proposed. It comprises two major components: (i) a Portuguese Sign Language dataset and (ii) a duo-interaction dataset, between Deaf and/or hearing people. The database can be used for different purposes like Sign Language recognition tasks or emotion/expressiveness recognition from body language. The whole database along with manual annotations, including signs and body parts, will be available for research and benchmark purposes.

1 Introduction

Sign Language (SL) is the medium of communication between hearing impaired people. It is a multimodal language, since it involves manual (i.e. hand gestures) and non-manual signs (i.e. facial expressions, head motion, pose and body movements). Therefore, emotions and expressions play a crucial role to convey meaning. As SL is commonly used just within deaf communities, with its own lexicon and grammar, the majority of hearing people are unfamiliar with the SL. In order to overcome the communication barrier between deaf and hearing people, several sign language recognition (SLR) systems have been proposed [1]. In addition, automated expressiveness/behavior recognition from body language can also be used to reduce the major gaps that currently prevent deaf people from easily interact with hearing people.

The evaluation and validation of such automatic recognition systems rely on the availability of appropriate ground-truth data. Herein, a video-based SL and body expressiveness database, called CorSiL, is presented. It can be used for the evaluation and validation of (i) SLR systems and (ii) expressiveness/behavior recognition systems. In this regard, the CorSiL database is composed by two distinct datasets, one suitable for SLR tasks, called signLangDB, and another for expressiveness recognition from body behavior, called corpLangDB. Both datasets have been already manually annotated. By the end of the annotation task, the entire CorSiL database will be made freely available to the research community for benchmark purposes. To our knowledge it will be the first database that gathers SL videos along with videos depicting the duo-interaction between deaf and/or hearing people. This composition makes the CorSiL database so unique and valuable, since the possibility of understanding the emotions and expressiveness behind the signs may open new research paths in SLR.

2 Related Work

Several sign language databases have been proposed in the literature. A brief review on video-based SLR databases is presented in [3]. The SL datasets can be roughly classified in two main groups: for the isolated sign recognition and the continuous sign recognition. A selection of the most relevant benchmark datasets in SLR is presented below.

The Purdue RVL-SLLL ASL Corpus [4] is an available database of the American Sign Language (ASL) suitable for both isolated and continuous SLR. The database provides a wide range of signed material, including 62 isolated gestures representing the numbers and the alphabet as well as examples of short discourse narratives. The database was collected from 14 signers under controlled lightning conditions in a uniform background. The RWTH-BOSTON Corpora [2] is composed by different subsets created to be used for benchmarking of ASL recognition systems. The acquisition and recording conditions were the same in all databases. Data was collected in a dark studio background and the signer's clothing was constrained with long sleeves. The signing data were recorded using four cameras: two cameras placed towards the signer's face for stereo vision purposes, one on the side and the other in the front, with close zoom on the face. The RWTH-BOSTON-50 Corpus, created for the task of isolated SLR, contains 50 signs that were performed by three signers. On the other hand, the RWTH-BOSTON-104 database, created for continuous SLR, comprises 201 continuous sentences constructed from 104 signs. For the evaluation of hand tracking methods, this database has been annotated with the signer's hand and head positions. The larger database of the RWTH-BOSTON Corpora, called RWTH-BOSTON-400, contains a total of 843 continuous sentences created from about 400 signs, performed by 4 native signers. Both Purdue and RWTH-BOSTON databases are not suitable to signer-independent continuous sign language recognition, since the Purdue database has a small number of sentences and the RWTH-BOSTON was only performed by 4 signers. A more complete database was presented in [5], the SIGNUM database. The database currently contains 450 basic signs, representing different words types, and 780 continuous sentences. The entire corpus was performed once by 25 native signers in a controlled environment along with clothing constrains.

Although several SL databases have been proposed in the literature, many issues remain unexplored: (i) there are no Portuguese Sign Language (PSL) databases available; (ii) most of the available datasets were recorded in a very controlled scenario; (iii) there are few sign language databases that gather RGB color data with depth information; and (iv) there are no databases with videos depicting the interaction between deaf and hearing people. The database presented in this paper attempts to address all of these problems.

3 CorSiL Description

The CorSiL database was created within the framework of a research project at INESC TEC, FEUP, Porto, Portugal, which aimed to develop both a video-based automatic SLR system and a expressiveness analysis method of the human body during the interaction between a pair of subjects. In this manner the CorSiL database has two major components (or subsets) each one with a specific purpose:

1. **[signLangDB]:** a Portuguese Sign Language video dataset.
2. **[corpLangDB]:** a duo-interaction video dataset between Deaf and/or hearing people.

Besides video content, both datasets have technical annotations that are still in development. By the end of the annotation, the entire database will be made freely available for research and benchmark purposes in order to establish the first PSL database as well as the first duo-interaction database between Deaf and hearing people.

The contact with the signers and volunteers of the recordings was obtained with a partnership with the *Escola EB2/3 Eugénio de Andrade*

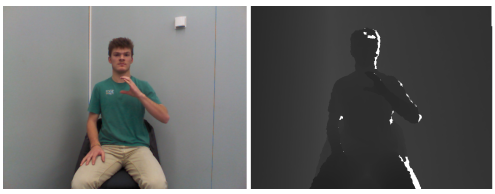


Figure 1: Color and depth pair of images from the signLangDB dataset.

and *Escola Artística de Soares dos Reis*, Porto, Portugal.

3.1 signLangDB subset

The signLangDB dataset is a PSL database suitable for both isolated and continuous SLR tasks. The dataset contains 182 isolated signs, representing the alphabet and the numbers as well as nouns, pronouns, verbs or common expressions, some performed with one hand and others with both. These signs include not only the informative part of the sign but also the entire movement from the rest position to the return to it. It also contains 40 continuous sentences that were selected in an attempt to comprise the most common situations that Deaf people might find in their daily life. All sentences are grammatically well-constructed in which there are no constraints regarding a specific sentence structure. In addition, no intentional pauses are placed between signs within a sentence.

All gestures and sentences were performed once by 15 native signers, including 5 males and 10 females, in a free and natural expression environment, without any clothing restriction but with a slightly-controlled uniform background. Moreover, some of the signers performed their gestures from a standing position while others performed seated in a chair. The recording conditions were set with this minimal amount of constrains so that they could meet a real environment scenario.

The signing data were acquired using the Microsoft Kinect camera, making this dataset one of the few with depth information associated to the RGB color data. The usage of depth eases the effort on preprocessing leaving space for heavier computational tasks. All videos were recorded using an image resolution of 640*x*480 at 30 fps. This spatial information should ensure a reliable extraction of hand and facial features from the images. Each video clip was stored as a sequence of *.png* images in order to speed up the access to individual frames. Figure 1 illustrates a pair of color and depth images.

The annotations of the signLangDB database include the segmentation of each sign and sentence for classification purposes as well as the identification of hand and head positions for tracking purposes.

3.2 corpLangDB subset

The corpLangDB dataset contains videos depicting the interaction that occur between a pair of individuals during a dialogue. The purpose of such a dataset is to enable the possibility of performing studies that analyse dialogue relationships (from sociological, psychological and technical perspectives) between two individuals, from distinct populations: Deaf and hearing people, in a relaxed environment. The conversation scenarios and topics recorded in this dataset were defined by socio-psychologists from the *Faculdade de Psicologia e Ciências da Educação da Universidade do Porto*. In this regard, the following three conversation scenarios were defined: 1) Conversation between two Deaf people; 2) Conversation between two hearing people; 3) Conversation between a Deaf and a hearing person.

In order to execute these scenarios, two requirements were defined so that the interaction between each pair of individuals could occur in the most natural way possible. Therefore, the individuals should know each other and have some kind of affinity and the acquisition should take place in a venue that was familiar to all subjects.

As the focus of the corpLangDB database is to enable the analysis of behaviour and expressiveness, the set of conversation topics was defined in a staggered way so that the discussion would generate emotions of increasing intensity in the actors of the conversation. To build a framework for the videos' acquisition, four different conversation topics were defined belonging to two-fold moments: positive (1 and 2) and negative (3 and 4): 1) Talk about happy moments; 2) Talk about people with which the actor has a strong love or friendship bond; 3) Talk about sad moments; 4) Talk about situations that awaken anger/indignation/injustice. The volunteer subjects, 13 in total, were coupled so that the conversation scenarios were covered. Each conversation between a pair of subjects was designated as a session having been a total of 9 sessions included so far in the database.

Figure 2: The same frame recorded from the cameras P0, P1, P2 and P3.

Figure 2 represents the entire scenario used to record the videos. IP0, IP1, IP2 and IP3 represent the cameras used and K a Microsoft Kinect. These were all placed in strategic locations (at a height of 2.58 meters) for the best capture possible. Two chairs were centred in the room in a way that was propitious for the dialogue in terms of proximity and comfort and for the video acquisition. The annotations available in this dataset include the positions of the head, hands, trunk, elbows, eyes, mouth, and nose. These annotations were performed using the VIPER-GT tool. [1]

4 Conclusions

This paper presents a video database, called CorSiL, that is composed by two distinct subsets, namely a SL dataset and a duo-interaction dataset (between Deaf and hearing people). The database includes technical annotations, such as temporal segmentation of the signs and body parts. The CorSiL dataset aims to provide a benchmarking tool for the comparison of both sign language recognition systems as well as expressiveness/behaviour analysis methods from body language. As future work, the technical annotations will be extended to the entire content of the database. In addition, further recordings will be conducted in order to increase the number of signers of the database.

5 Acknowledgments

This work is financed by the ERDF European Regional Development Fund through the COMPETE Programme (operational programme for competitiveness), FCOMP-01-0124-FEDER-037281 and by National Funds through the FCT Fundação para a Ciência e a Tecnologia (Portuguese Foundation for Science and Technology) within project PEst-C/EEI/LA0014/2013. The authors would like to thank both *Escola EB2/3 Eugénio de Andrade* and *Escola Artística de Soares dos Reis* for contacting the signers and providing the necessary means for this study. We also thank to Kelly Rodrigues and the Social Psychology Research Group of the University of Porto for their scientific advice. The fifth author would like to thank FCT for the financial support of the PhD grant with reference SFRH/BD/51430/2011.

References

[1] M.E. Al-Ahdal and N.M. Tahir. Review in sign language recognition systems. In *Computers Informatics (ISCI), IEEE Symposium on*, pages 52–57, 2012.

[2] P. Dreuw, C. Neidle, V. Athitsos, and *et al*. Benchmark databases for video-based automatic sign language recognition. *in LREC*, 2008.

[3] P. Dreuw, J. Forster, and H. Ney. Tracking benchmark databases for video-based sign language recognition. *in Trends and topics in computer vision*, 6553:286–297, 2012.

[4] AM. Martinez, R.B. Wilbur, R. Shay, and AC. Kak. Purdue rvl-slll asl database for automatic recognition of american sign language. In *Multimodal Interfaces, Fourth IEEE International Conference on*, pages 167–172, 2002.

[5] U. von Agris and K.-F. Kraiss. Towards a video corpus for signer-independent continuous sign language recognition. *Gesture in Human-Computer Interaction and Simulation*, 2007.

[1] `http://viper-toolkit.sourceforge.net/`

Efficient Supervised Relevance Criteria Tailored to Discrete Features

Artur J. Ferreira
aferreira@deetc.isel.ipl.pt

Instituto de Telecomunicações, and
Instituto Superior de Engenharia de Lisboa
Lisboa, PORTUGAL

Mário A. T. Figueiredo
mario.figueiredo@lx.it.pt

Instituto de Telecomunicações, and
Instituto Superior Técnico, Universidade de Lisboa
Lisboa, PORTUGAL

Abstract

The benefits of *feature discretization* (FD) techniques for machine learning and pattern recognition tasks are well-known. The use of FD leads to discrete-valued features with enough information for the learning task at hand, while ignoring minor fluctuations that may be irrelevant for that task. As a consequence, we obtain compact data representations for learning purposes, yielding both better accuracy and lower training time, as compared to the use of the original features. However, in many cases, mainly with medium and *high-dimensional* (HD) data, the large number of features usually implies that there is some redundancy among them. Thus, it is advantageous to apply *feature selection* (FS) techniques on the discrete features, keeping the most relevant ones, in order to improve the performance of machine learning tasks. In this paper, we propose relevance criteria for supervised FS techniques on discrete data, based on the histogram of the discrete feature. The experimental results on HD data show that the proposed criteria can achieve better accuracy than widely used relevance criteria, such as mutual information and the Fisher ratio.

1 Introduction

High-dimensional (HD) datasets (*i.e.*, with a large number of features) are becoming increasingly common in many different application domains of machine learning and pattern recognition (ML&PR), such as genomics, bioinformatics, computer vision, satellite image analysis, multimodal audio-visual processing, and others. When dealing with HD data in ML&PR tasks, one often resorts to *feature selection* (FS) and/or *feature discretization* (FD) pre-processing procedures. FS aims at finding an adequate subset of the original features, whereas FD achieves compact data representations, desirably ignoring irrelevant fluctuations on the data for the task at hand, leading to more robust classifiers. The FD methods that take into account feature interdependencies (the so-called dynamic methods) are usually preferable to their static counterparts, which discretize each feature individually. However, when dealing with HD data, dynamic FD methods have a prohibitive computational cost, and one has to resort to static methods. To overcome this computational burden, one can combine FS and FD techniques, yielding hybrid methods. The literature on both FS and FD is vast, with a wide range of unsupervised and supervised techniques. A comprehensive list of FS techniques can be found in [6]. Regarding FD, there are several comprehensive reviews, such as the recent survey in [5].

In this paper, we propose three supervised relevance criteria for FS, tailored to discrete data. After running some FD technique, we apply one of our relevance criteria in order to select and keep an adequate subset of features. The remainder of this paper is organized as follows. Section 2 briefly reviews existing supervised FS and FD techniques. Section 3 presents our relevance criteria for supervised FS. The experimental evaluation and their discussion are reported in Section 4.

2 Feature Selection and Feature Discretization

This Section briefly reviews some successful supervised FS and FD techniques, which have been proven successful for many problems. Regarding FS, we briefly review two relevance criteria widely used in supervised methods. The first one is the well-known *Fisher ratio* (FiR). In the multi-class case, $c > 2$, assuming class labels $j \in \{1, 2, \ldots, c\}$, the FiR for feature X_i [3] is

$$\mathrm{FiR}_i = \sum_{j=1}^{c} \mathrm{n}_j^{(y)} \left(\overline{X_i^{(j)}} - \overline{X_i} \right)^2 \Big/ \sum_{j=1}^{c} \mathrm{n}_j^{(y)} \mathrm{var}\left(X_i^{(j)} \right), \qquad (1)$$

where $\mathrm{n}_j^{(y)}$ is the number of occurrences of class j in the n-length class label vector y, and $\overline{X_i^{(j)}}$ denotes the sample mean of the values of X_i whose class label is j; finally, $\overline{X_i}$ is the sample mean of feature X_i.

Another relevance criterion widely used by FS methods is (the sample-based estimate of) its *mutual information* (MI) [2] with the class label. The MI between two *random variables* (RV) X and Y, is

$$MI(X;Y) = \sum_{i=1}^{N_x} \sum_{j=1}^{N_y} p_{X,Y}(i,j) \log_2 \left(\frac{p_{X,Y}(i,j)}{p_X(i) p_Y(j)} \right), \qquad (2)$$

in which $p_{X,Y}(i,j)$ is the joint probability of X and Y. MI is non-negative, being zero only when X and Y are statistically independent [2]. When using both FiR and MI for relevance-based FS on a d-dimensional dataset, we simply keep the $f \leq d$ top-rank features with the highest relevance.

Regarding FD, we briefly mention two successful supervised techniques: the well-known *information entropy minimization* (IEM) [4] and its variant, named *IEM variant* (IEMV) [7], which are based on the entropy minimization heuristic which yields a small number of discretization intervals. Both IEM and IEMV rely on information theoretical measures. In the literature of FD, we also find statistical-based and class-attribute methods; for more details see the survey [5].

3 Proposed Method

In this Section, we describe the proposed relevance criteria for discrete data and its usage in a FS filter. The relevance of the discrete feature is computed by checking the histogram of each discretization interval, across the different classes. The proposed method works as follows:

1. discretize all the d features in the dataset, by applying some FD method (e.g. one of the techniques mentioned in Section 2);
2. compute the *discretization levels/classes* (DLC) $l \times c$ matrix $\mathbf{D}$ for each feature (see Figure 1, for an illustrative example);
3. assess the relevance of each feature, by one of the criteria r_1, r_2, or r_3, described below; keep the most relevant features, according to the chosen criterion.

Figure 1 depicts the DLC matrices for two discrete features with four intervals ($l = 4$) and a three-class problem ($c = 3$) for a dataset with $n = 75$ patterns (25 per class). On the training data, for the patterns of each class, we compute the number of times that each discretization interval occurs for the patterns of each class. The relevance of the discrete feature is proportional to the non-uniformity of its histogram across the discretization levels and classes. For instance, for each row (discretization interval) of

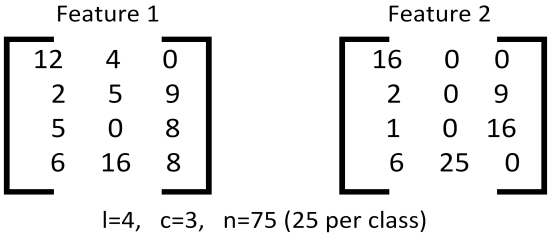


Figure 1: Discretization levels/classes matrix for two discrete features with four intervals (levels, $l = 4$) and a three-class problem ($c = 3$) with $n = 75$ instances (25 per class).

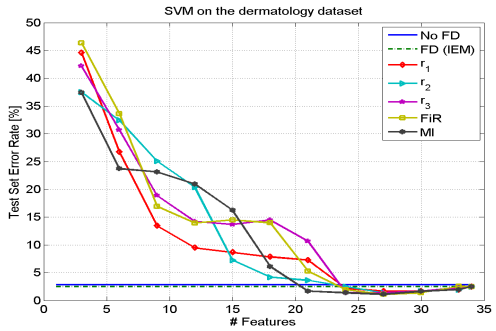


Figure 2: SVM test set error rate (%), 10-fold CV, for the Dermatology dataset ($d = 34$, $c = 6$, and $n = 358$), as functions of the number of features for different FS relevance criteria.

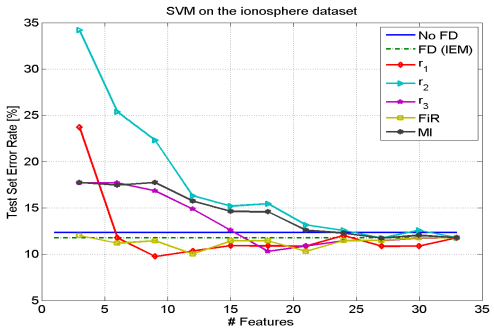


Figure 3: SVM test set error rate (%), 10-fold CV, for the Ionosphere dataset ($d = 33$, $c = 2$, and $n = 331$), as functions of the number of features for different FS relevance criteria.

the DLC matrix, an (almost) uniform distribution implies that such discretization level is not adequate to distinguish among all classes. However, if we get a non-uniform histogram it implies that we have quite different occurrences for patterns of the different classes. A special and interesting case is the occurrence of zeros in this matrix, which imply that a given discretization level never occurs in the patterns of a given class. Thus, in order to assess the relevance of feature X_i using the DLC matrix, we propose three relevance criteria, as follows:

- $r_1^{(i)}$ - the number of zero entries in the $\mathbf{D}$ matrix;
- $r_2^{(i)}$ - the sum of the absolute differences between the columns of the $\mathbf{D}$ matrix, given by

$$r_2^{(i)} = \sum_{k=1}^{c-1} \sum_{m=k+1}^{c} \|\mathbf{D}_k - \mathbf{D}_m\|_1 , \qquad (3)$$

  where $\mathbf{D}_k$ and $\mathbf{D}_m$ denote the k-th and the m-th columns of $\mathbf{D}$.

- $r_3^{(i)}$ - the trace of the matrix $\mathbf{A} = \mathbf{D_n}\mathbf{D_n^T}$, where $\mathbf{D_n}$ is the $\mathbf{D}$ matrix normalized such that its columns sum up one. The maximum trace of $\mathbf{A}$ is c and the higher the value of this trace the more the columns of $\mathbf{D}$ resemble the columns of the "identity matrix" (the ideal case).

These criteria can be applied to binary or multi-class problems. For the example in Fig. 1, these relevance values are: for feature 1, r_1=2, r_2=74, and r_3=1.14; for feature 2, r_1=5, r_2=132, and r_3=2.01. Thus, feature 2 will be considered more relevant than feature 1, which is in accordance with the intuitive analysis and the above considerations about the $\mathbf{D}$ matrix.

4 Experimental Evaluation and Discussion

We now report an experimental evaluation, carried out on public domain standard benchmark datasets, available from the UCI repository [1] or from the *gene expression model selector* (GEMS) repository[1]. We have considered the task of learning linear *support vector machine* (SVM) classifiers, using the Weka implementation[2], with its default parameters. The classification accuracy is assessed using 10-fold *cross validation* (CV). For each CV fold, the IEM FD method is applied to the training partition to obtain a quantizer, which is then applied to the test partition. We perform FS with our relevance measures and with MI and FiR.

Fig. 2 shows the test set error rate as functions of the number of features for the relevance-based FS filters using the following relevance measures: r_1, r_2, r_3, FiR given by (1), and MI given by (2), after IEM discretization. We use the Dermatology dataset which is a skin disease diagnosis problem, with $d = 34$ features, $c = 6$ classes, and $n = 358$ instances. The horizontal lines refer to the baseline error without FS; one of these lines corresponds to the absence of FD whereas the other corresponds to FD by the IEM method. The error rates reported by our three relevance criteria are competitive with those attained by FiR and MI.

Fig. 3 reports the results of a similar test of Fig. 2 for the Ionosphere dataset. The relevance given by both $r_1^{(i)}$ and $r_3^{(i)}$ produce adequate results, achieving better results than those attained by MI; the use of $r_1^{(i)}$ yields slightly better results than those of FiR. Table 1 reports the classification error attained by the same methods employed in both Fig. 2 and Fig. 3, for several datasets with a given number of features $f < d$. These results show that the proposed relevance measures attain results similar to those of FiR and MI, in different problems.

In summary, the proposed relevance criteria are useful for FS purposes, since they perform simple and efficient analysis on the histogram of the discrete features, aiming at finding the most discriminative ones. The classifiers learned on the selected features by our methods usually attain equal or better accuracy than those learned on the original discretized or non-discretized features. The proposed relevance criteria are competitive with two widely used relevance criteria for FS, on different problems, with r_2 and r_3 criteria achieving the best results. As future work, we will explore the use of these relevance criteria for FD.

References

[1] K. Bache and M. Lichman. UCI machine learning repository, 2013. URL `http://archive.ics.uci.edu/ml`.

[2] T. Cover and J. Thomas. *Elements of Information Theory*. John Wiley & Sons, 1991.

[3] R. Duda, P. Hart, and D. Stork. *Pattern Classification*. John Wiley & Sons, 2nd edition, 2001.

[4] U. Fayyad and K. Irani. Multi-Interval Discretization of Continuous-Valued Attributes for Classification Learning. In *Proc. Int. Joint Conf. Art. Intell. (IJCAI)*, pages 1022–1027, 1993.

[5] S. Garcia, J. Luengo, J. Saez, V. Lopez, and F. Herrera. A survey of discretization techniques: Taxonomy and empirical analysis in supervised learning. *IEEE Trans. on KDE*, 25(4):734–750, 2013.

[6] I. Guyon, S. Gunn, M. Nikravesh, and L. Zadeh (Editors). *Feature Extraction, Foundations and Applications*. Springer, 2006.

[7] I. Kononenko. On biases in estimating multi-valued attributes. In *Proc. Int. Joint Conf. Art. Intell. (IJCAI)*, pages 1034–1040, 1995.

[1] `www.gems-system.org`
[2] `www.cs.waikato.ac.nz/ml/weka`

Table 1: Test set error rate errors (%) attained by the SVM classifier, 10-fold CV, for datasets with c classes, n instances, sorted by increasing dimensionality d. We perform FS to select subsets with $f < d$ features. The best results (lower error) are shown in bold face.

			Feature Selection				
Dataset $(d;c;n)$	f	***FD (IEM)***	r_1	r_2	r_3	***FiR***	***MI***
Heart (13;2;270)	10	**12.9**	24.4	14.8	16.2	15.1	14.0
Wine (13;3;178)	10	1.6	1.6	3.3	**0.5**	2.2	2.7
Hepatitis (19;2;155)	14	18.7	24.5	20.7	**16.8**	16.9	19.4
Ionosphere (33;2;331)	9	11.7	**9.7**	22.3	16.8	11.4	17.7
Dermatology (34;6;358)	27	2.7	3.1	**1.6**	2.2	2.2	2.2
Colon (2000;2;62)	600	19.5	17.8	17.8	18.1	21.4	**16.1**
Brain-Tumor1 (5920;5;90)	1776	11.1	**8.8**	**8.8**	12.2	12.2	**8.8**
DLBCL (7070;2;77)	1092	5.0	6.4	**2.5**	6.2	5.0	5.1
ORL10P (10304;10;100)	3090	2.0	**0.0**	**0.0**	2.0	3.0	0.0
Brain-Tumor2 (10367;4;90)	4144	22.0	30.0	30.0	22.0	**20.0**	26.0
Prost.-Tumor (10510;2;102)	2100	**7.7**	**7.7**	**7.7**	**7.7**	**7.7**	**7.7**
Dexter (20000;2;2600)	7377	7.3	**6.3**	**6.3**	**6.3**	**6.3**	**6.3**

A comparison between Shannon's, Renyi's, and Tsallis Mutual Information for Feature Selection

Artur J. Ferreira
aferreira@deetc.isel.ipl.pt

Instituto de Telecomunicações, and
Instituto Superior de Engenharia de Lisboa
Lisboa, PORTUGAL

Mário A. T. Figueiredo
mario.figueiredo@lx.it.pt

Instituto de Telecomunicações, and
Instituto Superior Técnico, Universidade de Lisboa
Lisboa, PORTUGAL

Abstract

Feature selection (FS) techniques based on *mutual information* (MI) criteria have been used extensively in the last two decades. Since the seminal work of Battiti in the context of neural networks, many FS filters have been proposed. The results reported in the literature show the adequacy of these methods, for many different problems. However, the vast majority of these methods rely on information-theoretic measures proposed by Shannon. The extensions to Shannon's work as proposed, namely, by Renyi and Tsallis have received much less attention, in this context. This paper has a two-fold objective: explore the use of Renyi's and Tsallis definitions of MI in the context of FS for supervised classification tasks; bring some attention to the definitions proposed by Renyi and Tsallis. The reported experimental results on public domain benchmarks, show the adequacy of the three definitions for this purpose.

1 Introduction

Many machine learning and pattern recognition tasks often rely on the use of *feature selection* (FS) techniques, to attain adequate and compact subsets of features. FS may improve the accuracy of classification tasks while speeding up the training time. FS is a vast field of research; as a consequence, many unsupervised and supervised FS techniques have been proposed in the literature. A comprehensive listing of FS techniques proposed up to 2006 can be found in [7].

Many FS methods follow the *filter* approach [7], in which features are selected using some (relevance) criterion, without resorting to any learning methods. In most cases, with a large number of features, FS filters are the only applicable option, being preferable to embedded and wrapper methods [7]. FS filters usually apply a statistical or information-theoretical approach to rank or weight the features. The information-theoretical filters very often resort to the use of *mutual information* (MI) [5] to assess feature relevance and redundancy.

In this paper, we apply supervised FS filters based on the MI definitions resulting from the entropies defined by Shannon, Renyi, and Tsallis. We compare the use of these three measures for FS, in a supervised classification task. The remainder of this text is organized as follows. Section 2 briefly reviews the definitions of MI proposed by Shannon, Renyi, and Tsallis and their use for FS problems. Section 3 reports an experimental evaluation and its discussion on public domain benchmarks, using MI-based supervised FS methods, for a supervised classification task.

2 Background

This Section reviews the definitions of MI, by Shannon, Renyi, and Tsallis. An overview on the use of MI for FS purposes is also addressed.

2.1 Mutual Information by Shannon, Renyi, and Tsallis

The Shannon entropy of a discrete *random variable* (RV) $X \in \{1,...,N\}$, with each outcome having probability p_X, is defined as

$$H_s(X) = -\sum_{i=1}^{N} p_x(i) \log_2 p_x(i), \tag{1}$$

and satifies $0 \leq H_s(X) \leq \log_2(N)$ [5]. The entropy $H_s(X)$ is zero if one of the outcomes happens with probability one (no uncertainty) and attains its maximum value if X has a uniform probability mass function (maximum uncertainty). The MI associated to Shannon's entropy for two random variables $X \in \{1,...,N\}$ and $Y \in \{1,...,M\}$ is

$$MI_s(X;Y) = \sum_{i=1}^{N} \sum_{j=1}^{M} p_{X,Y}(i,j) \log_2 \left(\frac{p_{X,Y}(i,j)}{p_X(i)\, p_Y(j)} \right), \tag{2}$$

where $p_{X,Y}$ is the joint probability mass function of the pair (X,Y) and p_X and p_Y the corresponding marginals. MI is non-negative, and is zero if and only if X and Y are statistically independent [5]. At the other extreme, if X and Y are deterministically dependent, then $MI_s(X;Y) = \min\{H_s(X), H_s(Y)\}$. MI is related with the conditional entropy through the following well-known equalities: $MI_s(X;Y) = H_s(X) - H_s(X|Y) = H_s(Y) - H_s(Y|X)$ [5].

The parametric family of Renyi entropies is defined as

$$H_r^{\alpha}(X) = \frac{1}{1-\alpha} \log_2 \left(\sum_{i=1}^{N} p_X^{\alpha}(i) \right), \tag{3}$$

with $\alpha \geq 0$ and $\alpha \neq 1$. It can be shown that $\lim_{\alpha \to 1} H_r^{\alpha}(X) = H_s(X)$, that is, Renyi's entropy converges to the Shannon one. The special case of $\alpha = 2$ yields the so-called quadratic entropy. The MI corresponding to Renyi's entropy is

$$MI_r^{\alpha}(X;Y) = \frac{1}{\alpha-1} \log_2 \left(\sum_{i=1}^{N} \sum_{j=1}^{M} \frac{p_{X,Y}^{\alpha}(i,j)}{p_X^{\alpha-1}(i)\, p_Y^{\alpha-1}(j)} \right), \tag{4}$$

with $\alpha \neq 1$. Again, it can be shown that $\lim_{\alpha \to 1} MI_r^{\alpha} = MI_s$, as defined in (2). A detailed discussion and applications of the extensions proposed by Renyi can be found in [10].

Finally, Tsallis' entropy [11] is defined as

$$H_t^{\beta}(X) = \frac{1}{\beta-1} \left(1 - \sum_{i=1}^{N} p_X^{\beta}(i) \right), \tag{5}$$

where $\beta > 1$ is the entropic-index, and the corresponding MI is given by

$$MI_t^{\beta}(X;Y) = H_t^{\beta}(X) - H_t^{\beta}(X|Y) = H_t^{\beta}(Y) - H_t^{\beta}(Y|X). \tag{6}$$

More details on Tsallis entropies, related measures, and applications can be found in [11]. Both the Renyi and Tsallis entropies depend on a (free) parameter whereas Shannon's definition does not. Thus, we have more flexibility with Renyi's and Tsallis' definitions; of course, this flexibility brings the question of how to choose these parameters. In [9], the use of both H_r and H_t for the splitting criterion used in the well-known C4.5 decision tree learning algorithm is studied.

2.2 Feature Selection

Over the past two decades, many information-theoretic FS methods have been proposed [3]. Battiti was among the first to apply MI to FS, for the supervised training of neural networks [2]: the *MI-based feature selection* (MIFS) algorithm starts by computing the MI between each feature and the class label and chooses as the first feature the one that maximizes this measure. Then, it performs a forward iterative procedure that adds features that maximize a cost function which encompasses the MI between the candidate feature and the already selected features, with respect to the MI with the class label. For recent discussions on FS filters based on MI see [3, 8], and the many references therein. In [3], these methods are analyzed in a unified theoretical framework.

Recently [6], the potential inadequacy of MI for FS was addressed. Despite its popularity and appealing properties, the authors show that MI is not always the most appropriate criterion since searching for a feature subset that maximizes the MI with the class label does not always guarantee to decrease the generalization error. However, MI is still a (good) heuristic approach, with no guarantee in terms of classification accuracy.

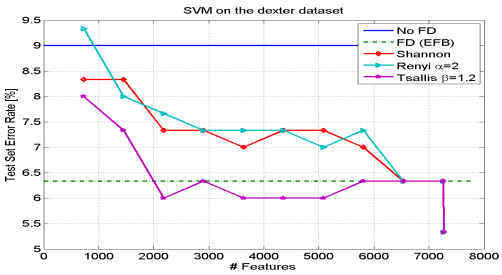


Figure 1: SVM test set error rate (%), 10-fold CV, for the Dexter dataset, as functions of the number of features.

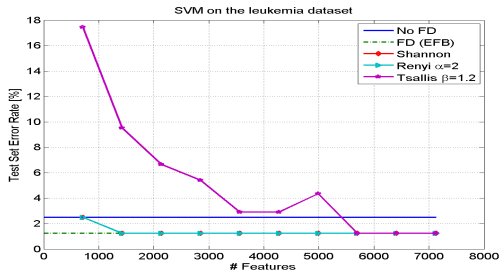


Figure 2: SVM test set error rate (%), 10-fold CV, for the Leukemia dataset, as functions of the number of features.

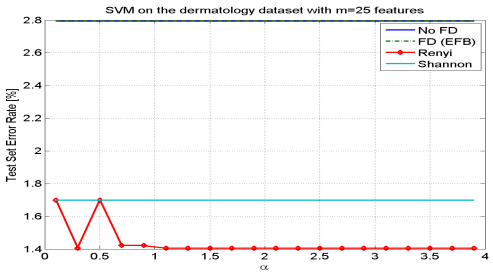


Figure 3: SVM test set error rate (%), 10-fold CV, for the Dermatology dataset, selecting subsets with $m = 25$ features, for MI_r with varying α.

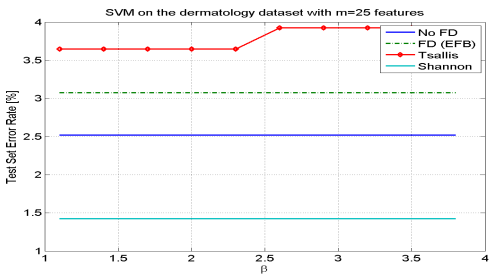


Figure 4: SVM test set error rate (%), 10-fold CV, for the Dermatology dataset, selecting subsets with $m = 25$ features, for MI_t with varying β.

3 Experimental Evaluation and Discussion

This Section reports an experimental evaluation of supervised FS methods, based on MI_s, MI_r, and MI_t. We compare these measures for supervised FS in a supervised classification task, with public domain standard benchmark datasets (from the UCI repository [1]). The task is to learn linear *support vector machine* (SVM) classifiers, using the Weka implementation `www.cs.waikato.ac.nz/ml/weka`, with its default parameters. The classification accuracy is assessed using 10-fold *cross validation* (CV). On each CV fold, prior to FS, we discretize each feature data with the unsupervised *equal-frequency binning* (EFB) [4], using 8 intervals.

We start with a comparison between the three MI definitions. Fig. 1 shows the test set error rate for the Dexter dataset (a text categorization problem) whereas Fig. 2 reports the results attained on the microarray dataset Leukemia. The error rates are displayed as functions of the number of selected features m, from a $d-$dimensional dataset; for each MI definition, we keep the m top-ranked features, with $m \leq d$. We have used $\alpha = 2$ (the commonly used quadratic entropy) and $\beta = 1.2$. The horizontal lines refer to the baseline error (on the original and discrete data), that is, without using FS; one of these lines corresponds to the absence of *feature discretization* (FD) whereas the other corresponds to EFB discretization. For both datasets, the error rates reported using MI_s and MI_r are similar. MI_t achieves better results on the sparse text categorization dataset than on dense microarray data.

We now asses the effect of changing the α and β parameters, when selecting a subset with $m < d$ features. Fig. 3 presents the results for different values of α on MI_r, whereas Fig. 4 shows a similar test changing β for MI_t. In both cases, we use the Dermatology dataset with $m = 25$. MI_r attains better results than MI_t, being able to achieve better results than MI_s, in a wide range of α values. For MI_t, these values of the β parameter never lead to an improvement on the test set error rate. Table 1 reports the classification error for several datasets, using $m < d$ features. We have chosen the value of m, such that $m \approx d/10$ or $m \approx d/5$. MI_s and MI_r attain comparable results, being better than MI_t, except for the (sparse) Dexter dataset. Both MI_s and MI_r yield a huge dimensionality reduction, keeping or improving the baseline error.

Table 1: Test set error rate errors (%) of the SVM classifier, 10-fold CV, for datasets with c classes and n instances, sorted by increasing dimensionality d. We perform FS to select a subset of $m < d$ features, with $\alpha = 2$ and $\beta = 1.2$. The best results (lower error) are shown in bold face.

Dataset $(d;c;n)$	m	***No FD***	***FD (EFB)***	MI_s	MI_r	MI_t
Brain-Tumor1 (5920;5;90)	592	**11.1**	14.4	12.2	13.3	16.6
Brain-Tumor2 (10367;4;90)	2072	22.0	20.0	**18.0**	20.0	30.0
Prost.-Tumor (10510;2;102)	1050	7.8	**6.8**	**6.8**	**6.8**	12.7
ORL10P (10304;10;100)	2060	2.0	2.0	**0.0**	**0.0**	7.0
Dexter (20000;2;2600)	2178	9.0	6.3	7.3	7.6	**6.0**

In summary, we have compared three different measures of mutual information, given by Shannon, Renyi, and Tsallis for the task of supervised feature selection. The experimental results, on public domain data, have shown that Shannon and Renyi definitions lead to similar results on dense data; in some cases, Renyi's definition is better. For sparse data, Tsallis definition achieves the best results. As future work, we will further investigate separately these measures on dense and sparse data, in order to narrow the possible range of the entropic indexes. We will also apply these measures on relevance-redundancy based feature selection filters.

References

[1] K. Bache and M. Lichman. UCI machine learning repository, 2013. URL `http://archive.ics.uci.edu/ml`.

[2] R. Battiti. Using mutual information for selecting features in supervised neural net learning. *IEEE Trans. N. Net.*, 5:537–550, 1994.

[3] G. Brown, A. Pocock, M. Zhao, and M. Luján. Conditional likelihood maximisation: a unifying framework for information theoretic feature selection. *JMLR*, 13:27–66, 2012.

[4] J. Catlett. On changing continuous attributes into ordered discrete attributes. In *European Working Session on Learning on Machine Learning (EWSL)*, pages 164–178. Springer, 1991.

[5] T. Cover and J. Thomas. *Elements of Information Theory*. John Wiley & Sons, 1991.

[6] B. Franay, G. Doquire, and M. Verleysen. Theoretical and empirical study on the potential inadequacy of mutual information for feature selection in classification. *Neurocomputing*, 112(0):64 – 78, 2013.

[7] I. Guyon, S. Gunn, M. Nikravesh, and L. Zadeh (Editors). *Feature Extraction, Foundations and Applications*. Springer, 2006.

[8] H. Liu, J. Sun, L. Liu, and H. Zhang. Feature selection with dynamic mutual information. *Pattern Recognition*, 42(7):1330 – 1339, 2009.

[9] T. Maszczyk and W. Duch. Comparison of Shannon, Renyi and Tsallis entropy used in decision trees. In *ICAISC 2008*, volume 5097 of *LNCS*, pages 643–651. Springer, 2008.

[10] J. Príncipe. *Information theoretic learning: Renyi's entropy and kernel perspectives*. Springer, 1st edition edition, 2010.

[11] C. Tsallis. Possible generalization of Boltzmann-Gibbs statistics. *Journal Statistical Physics*, 52(1-2):479–487, 1988.

On-Line Evaluation of Open-Ended Object Recognition System

Hamidreza Kasaei [2]
seyed.hamidreza@ua.pt
Miguel Oliveira[1]
mriem@ua.pt
Gi Hyun Lim [1]
lim@ua.pt
Luis Seabra Lopes [2]
lsl@ua.pt
A. M. Tomé [2]
ana@ua.pt

[1] IEETA
Universidade Aveiro
Aveiro,Portugal

[2] DETI/IEETA
Universidade Aveiro
Aveiro,Portugal

Abstract

In order to adapt to new environments, services robots should process visual information continuously, and perform learning and recognition simultaneously. This paper proposes the implementation of a simulated user to assess the performance of a 3D object recognition system in an open-ended setting. In this paper, "open-ended" implies that the set of object categories to be learned is not known in advance. With the proposed strategy the recognition system is able to interact with human users and learning new object categories continuously over time.

1 Introduction

3D object detection and recognition is increasingly used for manipulation and navigation tasks in service robots. Many recognition methods for both 2D and 3D data have been proposed however most of the proposed systems, training and testing are separated processes and do not occur simultaneously. Therefore, the systems are unable to adapt to dynamic environments. Open-ended strategy for object recognition implies that the set of object categories to be learned is not known in advance. There are some approaches which support incremental learning of object categories. Yeh and Darrell [6] developed novel methods for efficient incremental learning of SVM-based visual category classifiers, and showed that, using their framework, it is possible to adapt the classifiers incrementally. Kirstein et al. [5] suggested a life-long learning approach for interactive learning of multiple categories. They showed successful learning of 5 color categories and 10 shape categories observed in 56 objects. However it is followed a standard train and test strategy. In [3], [4] open-ended object category learning systems are proposed either based on one-class or multiple classifier schemes. Additionally, the authors also discuss the performance evaluation in open-ended schemes. The framework allowed the user to provide the names of objects through pointing and verbal teaching actions. The user could also ask questions about the categories of objects under shared attention and, if appropriate, provide corrective feedback for them. Therefore the authors also proposed the teaching protocol for evaluating the performance of open-ended object learning and recognition approaches.

In this paper, an interactive 3D object learning and recognition approach is presented. The approach is designed to be integrated in an open-ended learning system, which is used by an autonomous service robot working in a restaurant scenario. The experimental set-up concentrates on learning and recognizing table-top objects, which can be manipulated by the robot. This paper focuses on the development of a simulated user to assess the performance of the system in an open-ended setting.

2 Overall system architecture and on-line evaluation

In this paper, the problem of interactive open-ended object learning is addressed. In fact, it is impossible for a service robot to be completely programmed once it is started up. The robot must robustly adapt to new environments, either learning autonomously from experiences, or receiving feedback from human teachers. To cope with this, we proposed an open-ended learning approach for object recognition [1][2]. From a global perspective, the system, described in detail [1], is composed of several software modules such as object detector, feature extractor, object conceptualizer, user interface and object recognizer. The overall system architecture is depicted in Fig. 1.

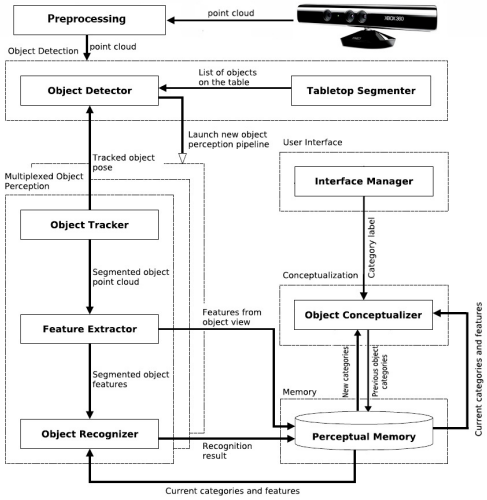


Figure 1: Overall architecture of the proposed system.

- The object detection module is responsible for detecting objects in the scene.
- The feature extractor module uses a 3D shape descriptor to compute regional features for the given object. This module computes spin-images corresponding to a fixed number of points of the view of the object point cloud. Spin-images are $2D$ histograms whose bins can have different width constituting then an user assigned parameter.
- The user interface module handles the labeling of the object.
- The object conceptualizer reads the current object categories from memory, as well as a set of features describing labeled objects. With this data, the object conceptualizer module creates or improves the description of an object category. The view might be stored as well as the Intra-Category distance (ICD) might be recalculated [1].
- The object recognizer module works in real-time. In recognition situations, a nearest-neighbor classification rule is used to assign category labels to the detected objects according to object templates stored in memory. However the decision rule additionally applies a threshold to the normalized object-to-category distances (distance of object to category divided by ICD), allowing to declare a target object as "unknown" if larger than a threshold (classification threshold).

On-line evaluations were conducted using the teaching protocol proposed in [3]. According to the teaching protocol, the system begins with zero knowledge and the training instances become gradually available, and the system incrementally builds object category models. Although the teaching protocol is designed to test the system based on categories that have been previously taught, the system can still reply that a given

object belongs to an unknown category. Therefore, the classification success must be evaluated in terms of precision and recall, leading to the use of F-measure as the measure of classification success. Average success is computed using a sliding window of size $3n$, where n is the number of categories that have already been introduced. If k, the number of iterations since the last time a new category was introduced, is less than $3n$, all results are used. In case the recognition fails to predict the correct category, both a false negative and a false positive are accounted for. Classification success is used as the indicator that a new category can be taught. According to the protocol, the system is ready to learn a new object category when the success is higher than a certain threshold, and at least one instance of every known category has been tested, i.e., $k \geq n$.

A simulated teacher is developed to follow the teaching protocol and autonomously interact with the system using *teach*, *ask* and *correct* actions. The simulated teacher repeatedly picks object views of the currently known categories from a database and presents them to the system for checking whether the system can recognize them. If not, the simulated teacher provides corrective feedback.

3 Experimental results

Two types of experiments, including *off-line evaluation* and *on-line evaluation*, were carried out to evaluate the proposed system. For these experiments, a restaurant object dataset was acquired, from 100 views of tabletop scenes by running the object detector and storing the segmented point clouds. These scenes contain 341 views of instances of 10 objects from different categories (bottle, bowl, flask, fork, knife, mug, plate, spoon, teapot, and vase) and 30 views of unknown objects. All tests were performed with an i7, 2.40GHz processor and 16GB RAM.

3.1 Off-line evaluation

To examine the performance of different configurations of the proposed system, an off-line quantitative evaluation based on leave-one-out cross validation strategy was used. Experiments were repeated for different values of three parameters of the system, namely the number of key points extracted from each object view, the image width of spin images and the classification threshold. A precision and recall for the 10 class problem above 0.8 was achieved using 10 key-points per object view, and computing the spin-images with 45 bins, and using 0.75 as classification threshold. These parameters were selected as default system configuration [1].

3.2 On-line evaluation

In this section, an on-line evaluation of the interactive open-ended object learning and recognizing is shown. The simulated user was used to evaluate the performances of the system. Initially, the system had no prior knowledge. In each run, the simulated user teaches a category and tests the system by presenting a previously unseen object view of the taught categories [3]. Throughout an experiment, the system improves the knowledge from the accumulation of experiences and the conceptualization of new object categories, and it must be able to recognize different object categories. When an experiment is carried out, learning performance is evaluated based on several measures, namely number of categories learned, number of question / correction iterations, average number of stored instances per category, global success (GS) and average classification success (ACS). Global success is the percentage of correct classifications made throughout the experiment. Average classification success is the average of all classification success values over all the question / correction iterations. Since the order of introduction of new categories may have a large impact on the performance of the system, four experiments were carried out that categories were introduced in random sequences. In all experiments, the system learned all 10 categories. Figure 2 shows the performance of the system throughout the experiments. By comparing all experiments (see Table 1), it is visible that in the first and last experiments the number of iterations required to learn 10 object categories was greater than other experiments. In the case of experiment 3, the success remained for the most part above the threshold. Results showed that both evaluation measures (global success and average classification success) for this experiment are higher than in all other cases. In addition, this was the experiment that learned all 10 categories faster.

Exp#	#Iterations	#Categories	#Instances	GS	ACS
1	63	10	4.3	0.80	0.88
2	57	10	4.1	0.83	0.89
3	54	10	3.5	0.90	0.92
4	63	10	4.6	0.77	0.83

Table 1: Summary of experiments

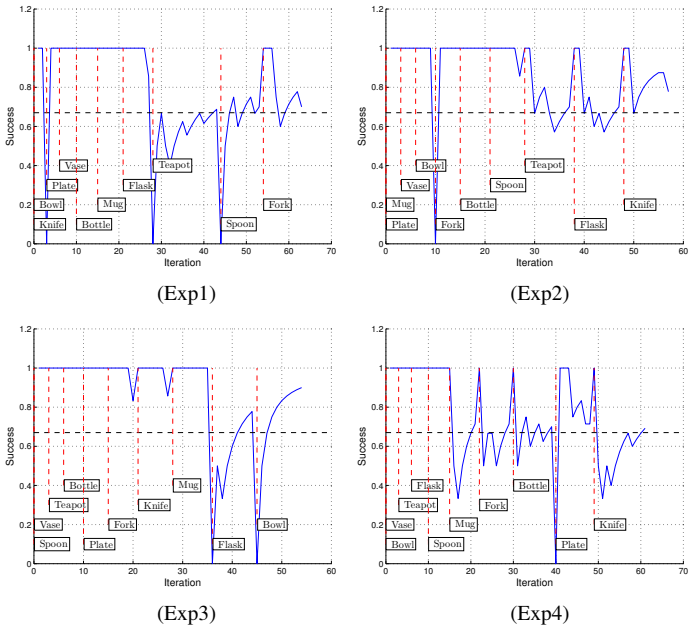


Figure 2: Evaluation of classification success versus number of iterations for different sequence of objects (success threshold set to 0.67, which marked by the horizontal line).

4 Concluding Remarks

A video of the proposed interactive open-ended learning for 3D object recognition is available at: http://youtu.be/XvnF2JMfhvc. The on-line evaluation proved that the system can incrementally learn new object categories. For future work, the database should be improved both in number of objects and number of categories. Furthermore, we would like to investigate other classifiers and object descriptors based on the bag-of-word concepts.

5 Acknowledgment

This research is partially funded by the RACE EU project FP7-287752, COMPETE/FEDER. The first author is supported by FCT under grant SFRH/BD/94183/2013.

References

[1] S. Hamidreza Kasaei, M. Oliveira, Gi Hyun Lim, L. Seabra Lopes, and A.M. Tome. An interactive open-ended learning approach for 3d object recognition. In *IEEE International Conference on Autonomous Robot Systems and Competitions (ICARSC)*, May 2014.

[2] Miguel Oliveira, Gi H Lim, Luis Seabra Lopes, Hamidreza Kasaei, Ana Tome, and Aneesh Chauhan. A perceptual memory system for grounding semantic representations in intelligent service robots. In *Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE, 2014.

[3] L. Seabra Lopes and A. Chauhan. How many words can my robot learn?: An approach and experiments with one-class learning. *Interaction Studies*, 8(1):53 – 81, 2007.

[4] L. Seabra Lopes and A. Chauhan. Open-ended category learning for language acquisition. *Connect. Sci*, 20(4):277–297, December 2008. ISSN 0954-0091.

[5] K. Stephan, H. Wersing, H. Gross, and E. Körner. A life-long learning vector quantization approach for interactive learning of multiple categories. *Neural Networks*, 28:90 – 105, 2012.

[6] T. Yeh and T. Darrell. Dynamic visual category learning. In *Computer Vision and Pattern Recognition, (CVPR 2008). IEEE Conference on*, pages 1–8, June 2008.

Segmentation and Tracking of the Mitral Heart Valve in 3D Echocardiography

Malik Saad Sultan
engr.saadsultan@gmail.com
Instituto de telecomunições,
Faculty of Science, University of Porto.

Miguel Coimbra
mcoimbra@dcc.fc.up.pt
Instituto de telecomunições,
Faculty of Science, University of Porto.

Abstract

Pathophysiological changes of the mitral valve demonstrate mitral valve prolapse/regurgitation and can be used for the early detection of pathological situations. An algorithm is proposed to automatically segment, track and finally estimate the leaflet morphology. Diagnostic decision and level of severity is based on measuring relevant mitral leaflet features such as thickness, leaflet mobility patterns, tenting volume and tenting height.

1 Introduction

Rheumatic heart disease (RHD) is the most common and main origin of heart valve disease (especially, mitral valve) in developing nations. In RHD, valve regurgitation is followed by valvular thickening and it directly affects leaflet mobility. Early detection is vital to minimize the risk of damage and to take preventive measures [1]. Diagnoses of mitral valve in ultrasound sequences is a challenging task due to fast, irregular leaflet motion, lack of reliable features and leaflet thickness in poor quality images [2].

Segmentation is required to locate the mitral valve leaflet in an ultrasound heart image. Researchers have proposed several semi-automatic and few fully automatic methods to segment and track the mitral leaflet. Transformation fitting is used to provide rough segmentation followed by two active contours for a more accurate identification of the valve and it requires manual initialization to start the algorithm [3, 4]. To be clinically feasible, it is very important to fully understand the complexities of patient-specific valve behaviour. Mitral valve segmentation and tracking has been well studied by Harvard Biorobotics Lab (HBL) group [5, 6, 7, 8]. They divide the segmentation problem into two parts. In the first part, the annulus has been segmented by using thin tissue detector and graph-cut method, by utilizing frame with closed valve. A constrained optical flow and valve state predictor is then efficiently utilized to track mitral annulus throughout the sequence. In a second step, a frame with an open valve is selected and the mitral leaflet has been found by using level set and graph-cut method. Recently, fully automatic algorithm segmentation methods are proposed by several researchers. Pouch et al.[9] efficiently captures patient-specific mitral valve leaflet geometry and his method integrates complementary probabilistic segmentation and shape modelling technique. Zhou et al.[2] and Liu et al.[10] proposed a method based on motion cues and the contiguous outliers detection in the low-rank representation. They analyze the fact that the motion patterns between the myocardium and the mitral leaflet are different, and thus motion cue can be used to track them.

Several methods have been proposed to estimate mitral valve prolapse and regurgitation. Accurate tracking of mitral valve make it feasible to analyze the anatomical structures of the mitral valve [11]. Mitral valve thickness has been associated with high risk of mitral regurgitation. The relation between leaflet thickness and motion pattern can be used to refine our results [12]. Mitral valve regurgitation is significantly correlated with the tenting volume and the tenting height [13]. Echocardiography is preferable to measure mitral valve properties and the relation between them because it is non-invasive, cheap, highly accurate and used for real-time assessment of the overall heart.

Our work is divided into two main parts. In the first part, algorithms are developed to automatically segment and track mitral valve. The second part of the work is focused on measuring mitral valve thickness, tenting volume, tenting height, motion pattern, geometry and the relation between them. Our work is aimed to increase diagnostic confidence for the automatic detection of RHD.

The rest of the paper is organized as follows. In Section 2, an algorithm is proposed for fully automatic segmentation and tracking of the mitral valve leaflet in 3D echocardiography. Once the mitral valve is located, a diagnostic algorithm is developed to estimate RHD problem and the level of severity. Results are presented in section 3 and finally the paper concludes with section 4.

2 Method

This paper aims to develop an unsupervised algorithm to automatically segment and track mitral valve throughout the cardiac cycle. The first part of algorithm is motivated by Zhou et al.[2] and Liu et al.[10] research work. In previous work, object tracking completely relies on the computation of optical flow which is unreliable and challenging under the conditions of low frame rate, very large deformations and fast motion (mitral valve motion). This paper uses the dynamic information throughout the cardiac cycle, relying on the fact that motion patterns between the myocardium and the mitral leaflet are totally different, feature points of the leaflet can be estimated by calculating cumulative energy distribution (CED) with respect to rank-K. Residue fitting of the valve region with fixed rank-k matrix is supposed to be much larger compared to the residue fitting of other regions. Therefore initial segmentation is based on low-rank matrix estimation, followed by active contours for shape refinement. Two regions are selected in a test sequence: the first region covers the mitral valve and the second region covers the myocardium and blood pool referred as the reference region (See Fig. 1). CED values of each region are then tracked and plotted with rank. Figure 1 shows that rank k=2 seems the most suitable choice to separate the selected regions. This approach provides a rough estimation of the mitral leaflet which is then refined by formulating the problem as contiguous outlier detection in the low rank matrix. Active contours may be used for shape refinement.

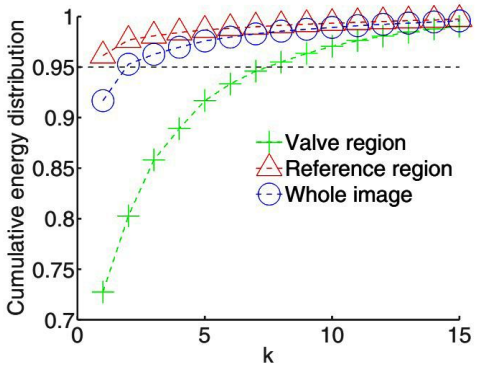


Figure 1: The graph of CED VS K provides a clear view of the difference of motion between the valve region and the reference region

Several criteria are defined to distinguish normal and diseased mitral valves (See Fig. 2). RHD may cause the leaflet narrowing or, in severe cases, result in the reversal of the blood flow. It happens because the leaflet becomes thickened, rigid and thus may stick together. Specifically for RHD, thickening of the leaflet is the most convincing criteria which provides a reliable feature to recognize a prolapsed mitral valve, which may cause regurgitation. There is a direct relation between mitral valve morphology and leaflet motion therefore leaflet thickening affects leaflet motion as well. However, it is observed that the increased leaflet thickness may not cause mitral valve regurgitation and thus might not be adequate criteria to detect severe cases. Mitral valve regurgitation can be analyzed by measuring tenting volume and tenting height. Seeing the fact we integrate tenting volume and tenting height to increase diagnosis confidence.

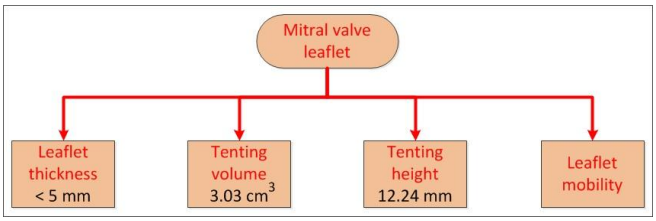


Figure 2: Criteria used to distinguish normal and diseased mitral valves (feature values presented are for normal mitral valves).

3 Result

The tracking algorithm is tested on several videos. Results demonstrate that the leaflet is continuously moving while the myocardium appears relatively stable. First row shows tracking result (See Fig. 3). Low rank matrix appears relatively stable and exclude mitral leaflet (See Fig. 3C). In this specific case the background sequence can be modelled as a low rank matrix. Tracking accuracy is shown by marking the correct mitral leaflet throughout the cardiac cycle (See Fig. 3B). Leaflet thickness is probably important to diagnose mitral valve prolapse but not as important to measure mitral valve regurgitation [12]. Therefore, tenting volume and height should need to be integrated with algorithm; idea is given in second row (See Fig. 3)

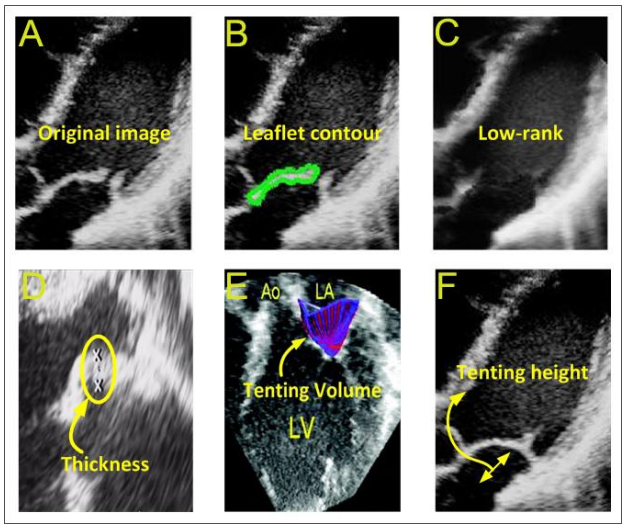


Figure 3: First row shows tracking results, whereas second row provide an overview of measured features.

4 Conclusion

In this paper efficient and fully automatic diagnostic system is proposed to measure mitral valve prolapse and regurgitation. The algorithm utilizes the difference in motion patterns to compute a rough segmentation of the mitral valve (locate the mitral valve) and then formulates the problem simply as a constrained outlier detection in the low rank matrix to accurately segment and track the mitral leaflet. Thus, this method automatically tracks leaflet motion without any user initialization and is an unsupervised approach. Literature shows that the measured thickness of the mitral valve indicates early mitral valve prolapsed and changes in leaflet motion due to thickness can be used to refine our decision ability. Tenting volume and height is deemed useful to measure severe situations which cause mitral regurgitation. Finally, this type of work can help end-users such as cardiologists, anaesthesiologist and surgeons to take better decisions during ultrasound procedures.

Acknowledgement

This work is funded by Instituto de telecomunições

References

[1] Eloi Marijon, David S. Celermajer, Muriel Tafflet. Rheumatic heart disease screening by echocardiograpghy. *American heart association*, pages 663–668, 2009.

[2] Xiaowei Zhou, Can Yang, Weichuan Yu. Automatic mitral leaflet tracking in echocardiography by outlier detection in low-rank representation. In Proc. *CVPR*, pages 972–979, 2012.

[3] S. Martin, V. Daanen, J. Troccaz, O. Chavanon. Tracking of the mitral valve leaflet in echocardiography images. 3rd IEEE international symposium on biomedical imaging: Nano to Marco, pages 181–184, 2006.

[4] S. Martin, V. Daanen, O. Chavanon, J. Troccaz. Fast segmentation of the mitral valve leaflet in echocardiography. Second international ECCV workshop, pages 225–235, 2006.

[5] Robert J. Schneider, Douglas P. Perrin, Nikolay V. Vasilyev. Mitral Annulus segmentation from 3D ultrasound using grapgh cuts. IEEE transaction on medical Imaging, Vol. 20, No. 9, pages 1679–1687, 2010.

[6] Robert J. Schneider, Neil A. Tenenholtz, Douglas P.Perrin, Gerald R. Matrix. Patient-specific mitral leaflet segmentation from 4D ultrasound. MICCAI, 2011.

[7] Robert J. Schneider, William C. Burke, Gerald R. Marx, Pedro J. del Nido. Modeling mitral valve leaflets from three-dimensional ultrasound. Functional imaging and modeling of the heart, Vol. 6666 of LNCS, pages 215–222, 2011.

[8] Robert J. Schneider, Douglas P. Perrin, Nikolay V. Vasilyev. Mitral annulus segmentation from three-dimensional ultrasound. IEEE international symposium on biomedical imaging, 2009.

[9] A.M. Punch, H. Wang, M. Takabe, B.M. Jackson, J.H. Gorman. Fully automatic segmentation of the mitral leaflets in 3D transesophageal echocardiographic images using multi-atlas joint label fusion and deformable medical modeling. Medical image analysis, Vol. 18, pages 118–129, 2014.

[10] Xin Liu, Yiu-ming Cheng, Shu-juan Peng, Qinmu Peng . Automatic mitral valve leaflet tracking in echocardiography via constrained outlier pursuit and region-scalable active contours. Neurocomputing, Vol. 144, pages 47–57, 2014.

[11] http://www.tomtec.de/home/best_place_to_work_2013.html

[12] Mario Senechai, Nicolas Michaud, Jimmy MacHaalany. Relation of mitral valve morphology and motion to mitral regurgitation severity in patients with mitral valve prolapsed. Cardiovasc ultrasound, 2012.

[13] J.M. Song, S. Fukuda, T. Kihara, M.S. Shin. Value of mitral valve tenting volume detertermined by real-time three-dimensional echocardiography in patients with functional mitral regurgitation. Am J cardiol, page 1088–1093, 2006.

Learning with Drift in Twitter

Joana Costa[12]
joana.costa@ipleria.pt,joanamc@dei.uc.pt
Catarina Silva[12]
catarina@ipleiria.pt,catarina@dei.uc.pt
Mário Antunes[13]
mario.antunes@ipleiria.pt,mantunes@dcc.fc.up.pt
Bernardete Ribeiro[2]
bribeiro@dei.uc.pt

[1] School of Technology and Management
Polytechnic Institute of Leiria, Portugal
[2] Center for Informatics and Systems
University of Coimbra, Portugal
[3] Center for Research in Advanced Computing Systems
INESC-TEC, University of Porto, Portugal

Abstract

Social networks have become part of our daily routine as Internet users. Reading news, looking for a service, asking for help, or simply sharing emotions or thoughts with family and friends turned social networks into a huge repository of information as users share daily valuable information. Learning in such a dynamic environment requires specific approaches, not only because of the diversity of data but because time plays an important role, drifting concepts over time. In this paper we propose a learning strategy to learn in the presence of concept drift in Twitter, one of the most well known social networks. Two learning models are proposed: a time-window model and an ensemble based model. We also present the QtSim framework, designed to simulate different types of drift by artificially timestamping real Twitter messages, that allows us to evaluate and validate our strategy. Results are so far encouraging regarding learning in the presence of drift, along with classifying messages in Twitter streams.

1 Introduction

Over the last few years, with the burst of social networks, people became easily connected and can communicate, share and join together. This can obviously endorse noteworthy changes in information spreading, as information is being shared publicly among users. One of the most well-known social media platforms is Twitter, a microblogging service where users post text-based messages, *tweets*, of up to 140 characters. Another interesting characteristic of Twitter is the presence of *hashtags*, single words started with the symbol "#", used to classify each message content. Along with the deluge of data created, time is an important constraint, as the flow of information is continuous and changes over time: one might be refering to an important event that might be occurring today, and in a few days those *tweets* might have disappeared and new content arises. Learning in the presence of concept drift is not an easy task and requires a specific approach. The learning model must have not only the ability to continuously learn, but also the ability to change concepts already acquired. To deal with concept drift in the Twitter stream we propose a two-fold approach: a time-window model and an ensemble based model. We also propose a framework to simulate different types of drift by artificially timestamping real Twitter messages in a sequential way in order to evaluate and validate our strategy. By studying different types of drift we aim to identify the learning characteristics best tailored to learn in such environments, where each drift might occur.

2 Related Work

In [1] an approach for *hashtag* recommendation in Twitter is introduced. This approach computes a similarity measure between *tweets* and uses a ranking system to recommend *hashtags* to new *tweets*. In [2] the use of *hashtags* to classify Twitter messages is done by clustering similar *tweets* in a graph based collective classification strategy. Although the presented results seem promising, we have identified the lack of adaptiveness in this strategy. A different approach is proposed in [3], where an event detection method is described to cluster Twitter *hashtags* based on semantic similarities between the *hashtags*. This work is in line with our previous work except for the fact that the semantic similarities are computed based on the message content similarities rather than being based on semantic *hashtag* similarities.

3 Proposed Approach

Twitter classification is a multi-class problem that can be cast as a time series of *tweets*. It consists of a continuous sequence of instances, in this case, Twitter messages, occurring each instance at a time, not necessarily in equally spaced time intervals, and is characterized by a set of features, usually words. A labelled instance is represented as a pair between the feature vector of that instance along with the associated class label.

We have used a classification strategy previously introduced in [4], where the Twitter message *hashtag* is used to label the message content. Notwithstanding the Twitter message classification is a multi-class problem in its essence, it can be decomposed in multiple binary tasks in a one-against-all binary classification strategy, which means one classifier for each class.

For classifying time series like the Twitter stream we propose a two-fold approach: a time-window model and an ensemble model. The time-window model is a batch learning model unable to retain all the previously seen examples. Differently, the ensemble model has a modular structure which enables temporal adaptation to new incoming *tweets* on the basis of the data sampling real distribution over time. The main purpose is to design a memory mechanism that allows newly seen examples to be identified based on past experiences.

Algorithm 1 defines the basic steps of the time-window model. For each collection of documents $\mathcal{T}$ in a time-window t, $\mathcal{T}^t = \{x_1,\ldots,x_{|T^t|}\}$ with labels $\{y_1,\ldots,y_{|T^t|}\} \rightarrow \{-1,1\}$, the dataset $\mathcal{D}^t$ is updated with the newly seen documents. No previously seen documents are stored in $\mathcal{D}^t$ and thus $\mathcal{C}^t$ classifier is always trained with the examples of the most recent time-window.

Algorithm 1: Time-Window Model

Input:
For each collection of documents $\mathcal{T}$ in a time-window t,
$\mathcal{T}^t = \{x_1,\ldots,x_{|T^t|}\}$ with labels $\{y_1,\ldots,y_{|T^t|}\} \rightarrow \{-1,1\}$ $\quad t = 1,2,\ldots$ T

1 **for** *t=1,2, ... T* **do**
2 $\quad \mathcal{D}^t \leftarrow \mathcal{T}^t$
3 **end**
4 BaseClassifier $\mathcal{C}^t$: Learn ($\mathcal{D}^t$), obtain: h^t: $\mathcal{X} \rightarrow \mathcal{Y}$ Time-Window
Classifier $\mathcal{C}^t$: Classify ($\mathcal{T}^{t+1}$), using: h^t: $\mathcal{X} \rightarrow \mathcal{Y}$

The ensemble model, presented in Algorithm 2, proposes to store all the information gathered with the previously seen examples. For each collection of documents $\mathcal{T}$, that contain both positive and negative examples and occur in a time-window t, a classifier $\mathcal{C}^t$ is trained and stored. When a new collection of documents in the subsequent time-window is presented to the ensemble model, all the previously trained classifiers are loaded, and each one will classify the newly seen examples. The prediction function of the ensemble, composed by the set of classifiers already created, is a combined function of the outputs of all the considered classifiers. Several strategies can be used herein. We propose a majority voting strategy where each classifier participates equally. When there is a tie, i.e. the votes account to zero, the classification of the most recent classifier is used to untie.

4 The QtSim Framework

In this work we have developed the QtSim framework that dynamically creates datasets by artificially timestamping real *tweets*. The major goal of this framework is to create labelled datasets that can be used to simulate different drift patterns that will evaluate and validate our previously introduced strategy. The framework receives a document set for each document class, typically *tweets* that contain the same *hashtag*, along with the frequency of that class during previously defined time-windows. The main idea is to use the frequency to reproduce artificial drifts. For instance, a sudden drift might be represented by *tweets* from a given *hashtag* that in a given temporal moment start to appear with a significant

Algorithm 2: Ensemble Model

Input:
For each collection of documents $\mathcal{T}$ in a time-window t,
$\mathcal{T}^t = \{x_1, \ldots, x_{|T^t|}\}$ with labels $\{y_1, \ldots, y_{|T^t|}\} \rightarrow \{-1, 1\}$ $\quad t = 1, 2, \ldots$ T

for *t=1, 2, … T* **do**
  $\mathcal{D}^t \leftarrow \mathcal{T}^t$
  BaseClassifier $\mathcal{C}^t$: Learn ($\mathcal{D}^t$), obtain: h^t: $\mathcal{X} \rightarrow \mathcal{Y}$
end
for *k=1, … , t* **do**
  ModuleClassifier $\mathcal{C}^k$: Classify ($\mathcal{T}^{t+1}$), using: h^k: $\mathcal{X} \rightarrow \mathcal{Y}$
end
Ensemble $\mathcal{E}^t$: Classify ($\mathcal{T}^{t+1}$), using:

$$e^t = \begin{cases} \frac{\sum_t h^t(\mathcal{T}^{t+1})}{|\sum_t h^t(\mathcal{T}^{t+1})|} & \text{if } \sum_t h^t(\mathcal{T}^{t+1}) \neq 0 \\ h^t(\mathcal{T}^{t+1}) & \text{if } \sum_t h^t(\mathcal{T}^{t+1}) = 0 \end{cases}$$

frequency. Besides artificially timestamping real tweets, our framework represents each *tweet* as a vector space model, also known as *Bag of Words*. In this representation the collection of features is built as the dictionary of unique terms present in the documents collections and each tweet is indexed with the *bag* of the terms occurring in it. We have also integrated in our framework the INDRI API from the Lemur Project (`http://www.lemurproject.org/`) to add more features like indexing, parsing and querying. As our main intent is to create datasets for text classification approaches our framework can also apply pre-processing methods like *stopword removal* and *stemming*. The framework creates datasets in the ARFF format and in SVMLight format.

5 Dataset

We have created a dataset using our QtSim framework in order to evaluate and validade our strategy. As previously stated, we used a classification strategy introduced in [4], where the Twitter message *hashtag* is used to label the message content. We have simulated 10 different drift patterns and are based on those proposed in [5] , namely (i) *sudden*, (ii) *gradual*, (iii) *incremental*, and (iv) *reoccurring*. We have represented 2 instances of sudden, gradual and incremental drifts, to represent both increasing (refered as #1) or decreasing (refered as #2) frequencies. Regularity is represented here to show *tweets* that occur in a continuous frequency, i.e. without drift. We chosen 10 different *hashtags*, one for each defined drift, representing mutually exclusive concepts and hence different classes, such as *realmadrid* and *literature*. Table 1 shows the chosen *hashtags* and the corresponding drift.

The Twitter API (`https://dev.Twitter.com`) was then used in October 2013 to request public *tweets* that contain the defined *hashtags*. Besides having requested more than 10.000 *tweets*, those containing no message content besides the *hashtag*, along with all in non-English languages were discarded. Finally, we used 5700 *tweets* that were split in 24 timewindows according to the drift patterns previouly defined. In each timewindow the number of *tweets* is variable, as for simulating the drift patterns each class frequency varies along with time it is not compensated by any other.

6 Results and analysis

Table 1 summarizes the results obtained considering the F_1 measure.

Analysing the table we can observe the time-window model scores 51.53% of F_1 and it is outperformed by the ensemble model with 60.31%. Besides performing better than the time-window in the majority of drifts, nevertheless, in the drift *Gradual #1* and in the drift *Incremental #1*, the ensemble scores 40.45% against 49.88% and 30.69% against 41.41%, respectively, which are significant results. These drifts have the particularity of being the only ones that increase their frequency over time, which seem to denote that their nature and the performance obtained are related. The explanation is that in the first occuring time-windows, the time-window models used in the ensemble tend to fail, as they have not seen enough positive examples. In the last time-windows they contribute equally to the output of the ensemble and influence in a negative way the classification provided by the ensemble. This does not occur in the drifts with a decreasing frequency because, as the frequency is decreasing, the newly created models have seen less positive examples, but when they start to influence the ensemble decision, that in the beginning is mainly composed by models that have seen much positive examples, the examples they have to identify are less (as the frequency is decreasing) and thus the ensemble fails in a smaller proportion.

	Hashtag	Time-window	Ensemble
Sudden #1	#bradpitt	55.93%	58.42%
Sudden #2	#realmadrid	60.22%	80.12%
Gradual #1	#ryanair	49.88%	40.45%
Gradual #2	#literature	45.08%	74.53%
Incremental #1	#twitter	41.41%	30.69%
Incremental #2	#ferrari	52.01%	61.72%
Reoccurring	#syria	73.59%	82.92%
Regular #1	#jobs	55.78%	55.53%
Regular #2	#sex	57.69%	88.05%
Regular #3	#nowplaying	23.71%	30.65%
	Average:	**51.53%**	**60.31%**

Table 1: Comparative results: F_1 measure

Moreover, in *Regular #1* the ensemble model is also outperformed, but in this case with less significant results, 55.53% against 55.78%. We believe that this is related to the tie mechanism, as the examples misclassified are just a few and are those in which there was a tie and the last model, that is called to untie, fails the decision. Finally it seemed strange in a first glance that *Regular #3* had such a bad performance, specially when compared with a pronunciated drift. The results might be explained by the *hashtag* we choose to represent it, *#nowplaying*. This *hashtag* is commonly used to refer songs that users are playing in their devices, and considering the spectrum of musics and artists we suspect that the diversity of those *tweets* compromises the performance of the classifier.

7 Conclusions

We have presented two models to learn in the presence of concept drift in Twitter streams: a time-window model and an ensemble based model. We have also presented the QtSim frameworks, used to simulate different types of drift by artificially timestamping real *tweets* to evaluate and validate our strategy.

The results obtained revealed the usefulness of keeping information already gathered and using different strategies in the awareness of different kinds of drift. More precisely, we have identified that the same learning model performs equally with drifts of the same nature, and that in the case of a decreasing frequency drift it is better to use a time-window model instead of an ensemble model. Another solution is to combine the ensemble so that models with less positive examples participate with less score than those better suited to identify positive examples. Though, as storing can be a constraint in the Twitter stream data, it is important in future approaches to identify an outdated example, and for how long it is useful to store examples. This can be done by analyzing different time-window sizes, so we can reach a balance between the computational burden of storing and processing and the usefulness of storing.

Future work will include a more intensive study of the drift patterns in Twitter in order to extend the learning models to include different weighting mechanisms in the ensemble model, as the models that compose the ensemble may contribute differently to the final decision in the presence of different drift patterns. Furthermore, another study is to identify if there are *tweets* more informative than others, so pruning strategies can be used to relief the computational burden.

Acknowledgment

We gratefully acknowledge iCIS project (CENTRO-07-ST24-FEDER - 002003).

References

[1] E. Zangerle, W. Gassler, and G. Specht. Recommending #-tags in twitter. In *Proceedings of the Workshop on Semantic Adaptive Social Web 2011 in connection with the 19th International Conference on User Modeling, Adaptation and Personalization, UMAP 2011*, pages 67-78, 2011.

[2] Y. Duan, F. Wei, M. Zhou, and H.-Y. Shum. Graph-based collective classification for tweets. In *Proceedings of the 21st ACM International Conference on Information and Knowledge Management (CIKM)*, pages 2323-2326, 2012.

[3] H. O. O Ozdikis, P Senku. Semantic Expansion of Hashtags for Enhanced Event Detection in Twitter. In *The First International Workshop on Online Social Systems (WOSS)*, 2012.

[4] Joana Costa, Catarina Silva, Mário Antunes, and Bernardete Ribeiro. Defining Semantic Meta-hashtags for Twitter Classification. In *Adaptive and Natural Computing Algorithms, Lecture Notes in Computer Science, vol. 7824. Springer Berlin Heidelberg*, pages 226-235, 2013.

[5] Indre Zliobaite. Learning under Concept Drift: an Overview. Tech. Report, Vilnius University, Faculty of Mathematics and Informatic, 2010.

Detecting Unknown Pattern Repeats: First Results Using a Probabilistic Approach

Raquel Sebastião
raquel.sebastiao@ua.pt
Armando J. Pinho
ap@ua.pt

Signal Processing Lab, IEETA / DETI
University of Aveiro
Campus Universitário de Santiago
3810-193 Aveiro, Portugal

Abstract

This paper addresses the problem of identifying repeated substrings in a large nucleotide sequence. Moreover, the proposed approach is able to find similar common substrings, as well as exact repeated ones. The problem is dealt by constructing an ensemble of decision trees: the estimated probability profile of a sequence is computed by choosing, at each symbol, the maximal conditional probability from the pool of constructed trees. The position of the similar repeated patterns in the sequence is found without constrains on the nature or shape of the pattern in any way.

1 Introduction

The most recent developments in bioinformatics science and technology have led to the sequencing of a large number of genomes. A genome contains the genetic information of any organism [8].Often, patterns or subsequences can be found recurrently along the whole genome. Such patterns can reappear consecutively or interspersed. When the pattern is repeated consecutively it is called tandem; if the reoccurrences are not adjacent it is called interspersed. If the pattern is exactly repeated (without any kind of variations), it is a perfect or exact repeat, whereas if the recurrences are slightly varied by mutations (such as insertions, deletions or point mutations) it is an approximate repeat.

Individual differences in DNA are like each individual's fingerprints. Also, it is known that the genomes of two related species share similar genes. Indeed, it is reasonable to assume that genes' mutations between two species are possible, without compromising the similarity of the functions carried out. Such mutations may occur inter or intra species. Moreover, multiple exact and approximate repeats have been related with human genetic diseases.

Along with this, it is expectable that repetitions in genome sequences should be approximate rather than exact. Despite the efforts devoted to develop computer algorithms to enhance research in genomic biology, the problem of efficient blind repetition finding is still important and challenging. Therefore, the focus of this paper is in the problem of finding approximate interspersed patterns, without constraining the nature and shape of the patterns in any way, neither any insight on the same, within the same sequence as well as in related sequences.

2 The Repeated Patterns Problem

A DNA sequence can be computationally represented as a string S of symbols from the nucleotide alphabet $\sum = \{A,C,G,T\}$ [4]. Considering that X is a substring of the sequence S and that x is the repeated pattern. According to the computational biology literature, repeated patterns can be classified as follows:

- An **Exact or Perfect Tandem Repeat (ETR)** is a substring X composed by adjacent concatenations of a pattern x perfectly repeated, such as: $X = xx\ldots x$. For example, consider the pattern $x = babbler$, the substring $X =$ **babbler***babbler* is an ETR.

- A **k - Approximate Tandem Repeat (ATR)** is a substring X composed by adjacent concatenations of a pattern x repeated with slight variations, such as: $X = x_1x_2\ldots x_n$ where each $x_i, i = 1,\ldots,n$ is, according to some distance metric, at most, at k-distance from the pattern x. Considering the pattern $x = babbler$, the substring $X =$ **babbler***babbles****bubbles*** is an example of an ATR.

- A **k - Exact Interspersed Repeat (EIR) is)** is a substring X composed by non-adjacent concatenations of a pattern x perfectly repeated, such as: $X = xs_1xs_2\ldots s_{n-1}x$,where each $s_i, i = 1,\ldots,n-1$ are distinct substrings. For example, consider the pattern $x = babbler$, the substring $X =$ **babbler***test***babbler***example***babbler** is an EIR.

- A **k - Approximate Interspersed Repeat (AIR) is)** is a substring X composed by non-adjacent concatenations of a pattern x repeated with slight variations, such as: $X = x_1s_1x_2s_2\ldots s_{n-1}x_n$, where x_i and s_i are defined above. Considering the pattern $x = babbler$, the substring $X =$ **babbler***test****bubbles****example****gabbled*** is an example of an ATR.

Repeated patterns can be distinguished between Single, if the substring X is the concatenation of two copies of the pattern x, and Multiple, which contains more than 2 copies. Biology distinguishes repeated patterns according to the length of the pattern x into micro-satellites (if x is composed by less than 6 base-pairs), minisatellites (if x is composed from 7 to 100 base-pairs) and satellites (if x is composed by more than 100 base-pairs) [1].

The proposed approach is designed to deal with Exact and Approximate Interspersed Repeats, without any constraint about the patterns neither any insight on repeated pattern.

3 Related Work

Recent years have seen an increase of concerns and research efforts devoted to DNA similarity problems [2, 5, 6, 7, 9]. The advances in genomic technology and the developments in computational analysis to handle huge amounts of data have given their contributions.

Recently, in [6, 7] the authors proposed the concept of "information profile" to blind discover genomic regularities. The idea relies on the fact that locations with low information are associated with the presence of repetitive sequences. Using compression methods, quantitative information of the local complexity of a DNA sequence is also provided.

In [5] it is addressed the problem of identifying conserved substrings occurring in homologous genes in a large number of sequences, based on truncated suffix trees to represent the sequences. In [9] the authors presented an efficient algorithm to find all approximate tandem repeats within a sequence, allowing at most k insertions, deletions and mismatches in each repeat. However, both approaches can not deal with blind repetition finding problems.

4 DT2DIR – Decision Trees to Detect Interspersed Repeats

To address the problem of finding unknown interspersed repeated patterns in DNA sequences, we propose the use of decision trees. A decision tree relies on the construction of a classification or a regression model in the form of a tree structure, which predicts the value of a target variable based on several attribute variables (input). It can be interpreted as a direct acyclic graph where each node is either a decision node with two or more successors or a leaf node [3].

A decision or internal node (non-leaf) has some condition based on attribute values (split-test) and a leaf node is labeled according to the mode of the classes that reach this leaf. In a classification tree construction, each example traverses the nodes of the tree from the root until it reaches a leaf (testing the appropriate attribute at each node and following the branch corresponding to the attribute's value in the example) and the class assigned to the "destination" leaf is used to classify the example.

In the case of exact repeated patterns, the proposed method relies on computing a classification tree. When dealing with approximate repeated patterns it exploits the advantage of using several trees rather than one.

Consider a DNA sequence, represented by the string S of symbols from the nucleotide alphabet $\sum = \{A,C,G,T\}$. At each symbol, the proposed method considers a sliding window of length L which contains the

attribute variables. Since in DNA problems the size of the alphabet is 4, this is the number of classes we are dealing with. In this approach, the target variable is the first symbol that lies out of the sliding window, in such a way that in the next step it will be included in the sliding window becoming an attribute variable. Within these settings, a classification tree is computed and, for each leaf, the probability distribution over the classes is estimated. These estimates are computed taking into account a conditioning context with finite and fixed size $L > 0$ composed by the elements inside the sliding window. Since the size of the alphabet is 4, the number of conditioning states is 4^L, which is the total number of combinations among the different classes of the attribute variables.

A sliding window (SW) of length L at example x_i is defined as: $SW = \{x_l : l = i-L+1,\ldots,i\} = \{x_{i-L+1\ldots i}\}$. Therefore, the probability of predicting the occurrence of the symbol s having the sliding window $x_{i-L+1\ldots i}$ as the conditioning context, $P(X_{i+1} = s|x_{i-L+1\ldots i}), \forall s \in \sum$, depends on the number of times that the conditioning context is followed by the symbol s. This means that repeated patterns are easier to classify. When computing the estimated probability profile of a given sequence, repeated patterns are associated with higher probabilities.

This approach works to detect exact interspersed repeats: all the elements inside the conditioning contexts are considered to compute a classification tree. Afterwards, at each step, the probability of predicting the symbol that is associated with each conditioning context is estimated using the constructed classification tree. Thus, the estimated probability profile of the whole sequence can be computed.

When dealing with approximate interspersed repeats, the classification results can be improved by considering several classification trees instead of one. For a sliding window of size L, the proposed approach computes L classification trees, feeding each one with $L-1$ attributes. Each tree provides, based on the conditioning context, the estimated probability assigned at each class (each one of the 4 symbols that compose the alphabet). Thereafter, the estimated probability profile is computed by choosing, at each symbol, the maximal conditional probability among the computed trees. Thus, the attribute that is left out when computing each one of the L trees is the symbol that is not considered and will be responsible for approximating the pattern.

5 Results on a Toy Example

The proposed method to find approximate repeats is evaluated using a toy example of a DNA sequence with approximated interspersed repeats. For this purpose, data with interspersed repeats were simulated. Varying the seeds, 10 different DNA sequences with 20k pair bases were generated. An exact block with 500 bases were copied and 3 times pasted along the sequence. Herewith, at locations 2500, 7500, 12500 and 17500 we have repeated patterns. Moreover, at each of the pasted copy we enforce 5 different point mutations.

At this preliminary stage and attending that repeated patterns present higher probabilities, such regions are found by visual inspection of the estimated probability profile. Figure 1 shows the polar representation of the logarithm of the estimated probability profile (after applying a moving average low-pass filter). The estimated probability profile was computed by an ensemble of 5 classification trees and considering a window with length of 10 bases. The 4 existing approximated interspersed repeats can be easily identified as stressed out by the letters and the thick lines.

6 Conclusions and Further Research

This paper presents a probabilistic approach to detect unknown pattern repeats in DNA sequences. Through classification trees, the proposed method is able to find exact and approximate interspersed repeats. Classification trees are used to compute the estimated probability profile of a sequence and repeats are found through visual inspection of regions with higher probabilities.

The presented results are encouraging enough to continue enhancing the approach proposed. Therefore, beyond the assessment of the ability to detect repeated patterns, computational time must be evaluated and analyzed. Also, the number of false alarms must be taken into account and properly addressed. A comparison with other state of the art methods must also be performed.

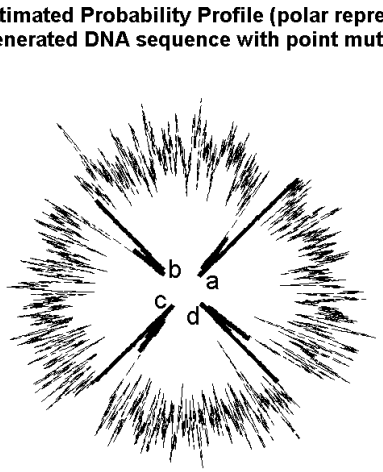


Figure 1: Polar representation of the log estimated probability profile of a generated DNA sequence (after applying a moving average low-pass filter). The letters and the thick lines indicate the detected repeated patterns.

It must be stressed out that no prior information about the size or the shape of the pattern which is repeated is needed. The identification is truly blind. The proposed approach is virtually capable of detecting repeats with any period, patterns repeated several times and patterns that are approximated by insertions and deletions. Nevertheless, further developments must take into consideration the evaluation and comparison of the ability to detect repeated patterns with different lengths.

7 Acknowledgements

This work was funded by project Cloud Thinking (funded by the QREN Mais Centro program, ref. CENTRO-07-ST24-FEDER-002031) and by National Funds through FCT - Foundation for Science and Technology, in the context of the project PEst-OE/EEI/UI0127/2014.

References

[1] Olivier Delgrange and Eric Rivals. Star: an algorithm to search for tandem approximate repeats. *Bioinformatics*, 20(16):2812–2820, 2004.

[2] Paulo Jorge S. G. Ferreira and Armando J. Pinho. Compression-based normal similarity measures for DNA sequences. In *IEEE International Conference on Acoustics, Speech and Signal Processing, ICASSP 2014, Florence, Italy, May 4-9, 2014*, pages 419–423, 2014. doi: 10.1109/ICASSP.2014.6853630.

[3] João Gama, Ricardo Fernandes, and Ricardo Rocha. Decision trees for mining data streams. *Intell. Data Anal.*, 10(1):23–45, January 2006. ISSN 1088-467X.

[4] Atheer A. Matroud, M. D. Hendy, and C. P. Tuffley. Ntrfinder: a software tool to find nested tandem repeats. *Nucleic Acids Research*, 40(3):e17, 2012.

[5] Ruby L. V. Moritz, Matthias Bernt, and Martin Middendorf. Local similarity search to find gene indicators in mitochondrial genomes. *Biology*, 3(1):220–242, 2014. ISSN 2079-7737. doi: 10.3390/biology3010220.

[6] Armando J. Pinho, Sara Garcia, Diogo Pratas, and Paulo Jorge S. G. Ferreira. Dna sequences at a glance. *PLoS ONE*, 8(11):e79922, 2013. doi: 10.1371/journal.pone.0079922.

[7] Armando J. Pinho, Diogo Pratas, and Paulo Jorge S. G. Ferreira. Information profiles for DNA pattern discovery. In *Data Compression Conference, DCC 2014, Snowbird, UT, USA, 26-28 March, 2014*, 2014. doi: 10.1109/DCC.2014.54.

[8] Matt Ridley. *Genome*. New York, NY: Harper Perennial, 2006.

[9] Dina Sokol and Justin Tojeira. Speeding up the detection of tandem repeats over the edit distance. *Theor. Comput. Sci.*, 525:103–110, 2014. doi: 10.1016/j.tcs.2013.04.021.

Non-rigid Registration of Point Clouds Regularized by NURBS

Hooshiar Zolfagharnasab
dee12012@fe.up.pt

Faculdade de Engenharia, Universidade do Porto
Rua Doutor Roberto Frias, S/N, 4200-465
Porto, Portugal

Hélder P. Oliveira
helder.f.oliveira@inescporto.pt

INESC TEC
Campus da FEUP, Rua Doutor Roberto Frias, 4200-465
Porto, Portugal

Abstract

Development of three dimensional (3D) modeling technologies has being widely used in medical application. Many different methodologies have been proposed in the past; however, low cost and easy-to-use equipment is highly desirable for the medical community. As a low-cost sensor, Microsoft Kinect has been adopted in many reconstruction methods; nevertheless, most of them fail to represent fully-aligned model due to non-rigidity between captured views. In this paper we propose a simple method to perform non-rigid registration, refining the result obtained from a rigid methodology. The implemented methodology was tested using breast cancer patients data.

1 Introduction

Within recent years, 3D modeling has enhanced medical scanning techniques to provide adequate information for specialists. Many methods have been recently proposed to use either 3D views or complete 3D models, in order to obtain relevant medical features; as in [5], to perform breast aesthetic evaluation after Breast Cancer Conservative Treatment (BCCT), authors extracted complimentary features from 3D frontal view of patients. Providing accurate 3D models to be used in medical applications, often requires expensive devices which are not available in every clinics. On the other hand, typically, scanning procedures are difficult to be performed by both technicians and patients, obtaining undesired non-rigid movement between views, which reinforces the importance to use low-cost and easy-to-perform devices.

Besides scanning issues, there are some difficulties affecting the 3D reconstruction of the different views. To cover that, a recent method was proposed to reconstruct patient model by using a 2-step rigid registration [1]; though reconstruction fails when scanned views have significant non-linear differences, due to non-rigid movements of human body. In this paper, we developed a methodology to improve the aforementioned rigid reconstruction [1], by fitting the views using a non-linear registration method, supported by a 3D structure using Non-Uniform Rational Basis Splines (NURBS).

2 Related work

2.1 Registration

3D models of human body has been studied widely in recent years [4], using low-cost devices, such as Microsoft Kinect, which provide a synchronized RGB and Depth information in each image pixel. Also colored point clouds can be obtained using Kinect output information. Although Kinect has been launched for video gaming applications, it was shown in [5] that the device is accurate enough to be used as a medical sensor to capture breast cancer patients' views, to be used for aesthetic quantification.

Since a single view does not have all necessary 3D information, registration methods are required to fuse different views into a complete 3D point cloud. Recently, applied on breast cancer patient images, a 2-step rigid registration method has been proposed [1, 2], based on an initial coarse registration followed by a fine registration stage. In the coarse registration step, views are approached to each other based on corresponding features. Such features can be either a point, a subset of points or shared characteristics between points of different views [2]. In this way, breast reconstruction reveals special difficulties because of featureless nature of human torso. Well-known coarse registration methods, using Principle Component Analysis (PCA) together with different Keypoint selection methods (i.e spin image, point curvature) fail or carry out poor reconstructed models since they cannot detect correct correspondences in different views [2]. In such manner Costa *et al*. [1] proposed to extract the global structure of the object in each view using Delaunay Triangulation method, as keypoint selection, and then align them using a rigid approach. The registration process is then applied by computing the rotation and translation, from the matched features between corresponding views. Finally, a fine registration step, using Iterative Closest Points (ICP) is used for fine adjustment of the views.

Although Costa's method [1] outperforms other coarse registration methodologies, namely in situations of high translation and rotation, when non-linear movements are present, the obtained reconstruction is not accurate, hence the method uses rigid transformation, and cannot deal with non-linear movements. In this manner, we propose a methodology to improve Costa *et al*. [1], using a non-linear transformation, supported by a 3D structure using NURBS.

2.2 NURBS

NURBS is widely employed in parametric fitting of curves and surfaces. Presenting stability, local modification and flexibility, it is supported in many modern 3D standards such as OpenGL and IGENS. Nevertheless, the quality of generated surface (or curve) is dependent to the structure of points which control the NURBS. Since almost all registration methods produce scattered and unorganized point clouds, in order to use NURBS structure in a registration stage, it is essential to organize them prior the surface generation [3].

Leal *et al*. [3] proposed to organize scattered points using a simple method. Initially, points are mapped into a plane found using PCA. Based on the projection of the points onto the regression plane of the cloud, the point cloud is regularized by inserting and removing points, respectively if in some regions there is lack of points, or if the cloud is too dense, creating an uniform grid. Such grid is used to create an order required to generate the NURBS. Refined points are then re-projected into point cloud while preserving their order. Finally, point cloud is optimized by modifying the weights of each point. The resulted organized point cloud will be fed to NURBS generator to draw the required surface.

3 Non-Rigid Registration

Although NURBS is not originally introduced to carry out registration, the idea behind this work is to use its structure in order to fit different point clouds. Previously discussed, rigid registration methods cannot handle non-rigid movements, since they are unable to manipulate individual points; however NURBS provides the necessary strategy to manipulate a point cloud by updating the location of a patch of points [6]. The registration pipeline discussed in [1] certifies that the views are well aligned; however some misalignment can be observed in the presence of non-linear movement. Proposed method refines points or patch of points, where this misalignment is detected.

Refinement of a view takes place by organizing its point cloud in a grid proposed by Leal *et al*. [3]. By preserving a control point of each grid (or cp), the organizing method simplifies and arranges the point cloud to be drawn as a NURBS. Next step is followed by NURBS generator which interpolates points between control points to generate a smooth surface. All the points are then grouped in different patches regarding the nearest control point. Considering point cloud P_i to be refined and point cloud P_j as reference, control point of each patch from P_i is updated as follows:

$$cp_k^i = average(p_{(x,y,z)}, q_{(x,y,z)}); \forall p \in P_i, q \in P_j : \|p - q\| > \delta$$

where p and q are corresponding patches from two point clouds, cp_i^k is control point of patch k on refining point cloud and δ represents the maximum allowed distance between patches, to be considered that a specific

patch is well registered with the reference. The method updates the control point of patch k to the average of the points of both patches on two point cloud, if the distance between patch k and the corresponding patch on the reference point cloud is greater than δ.

4 Experimental Results

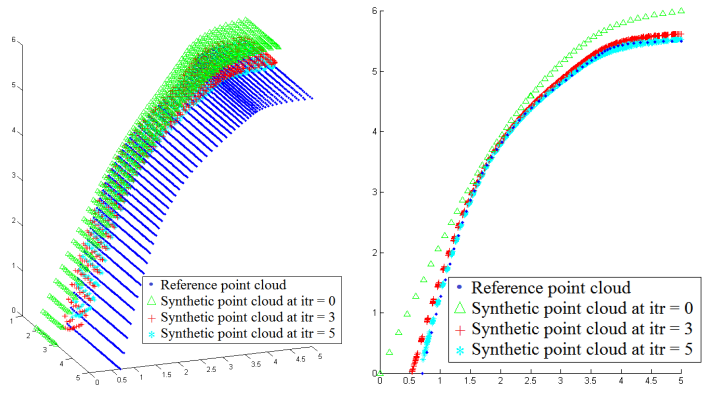
The proposed methodology was implemented in Mathworks MATLAB R2014a and executed on a machine with 8 GB of memory which was powered by Intel Corei7 vPro.

4.1 Dataset

Two experiments were conducted using synthetic data and scanned patient model. Synthetic data contained two views belonging to a surface that has some intersected areas. To test the ability of method to register them non-rigidly, point clouds of each view are not completely aligned. The other dataset contained two point clouds of the arm of a patient which were scanned using Kinect and rigidly registered by [1].

4.2 Results and Discussion

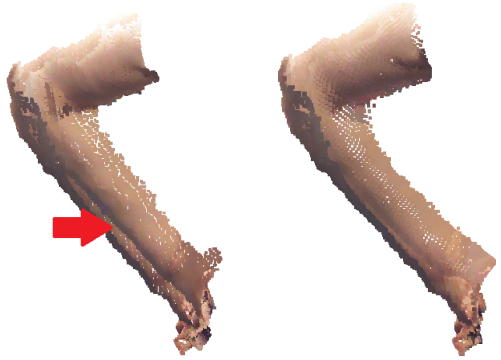
Synthetic data was used in order to demonstrate how the implemented methodology deal with non-linear misalignments. In Figure 1 we can see that the original point cloud is, iteratively, well registered with the reference in 5 iterations. Within each iteration, patches on synthetic point cloud are being approached to the corresponding patches on the reference point cloud, since control points of the patches are pushed to the reference during each iteration.



(a) (b)

Figure 1: Non-rigid registration of point cloud (△) onto the reference (.).

Additional experiment was conducted using data from breast cancer patient, containing the arm, in the presence of non-linear movement between views. In Figure 2(a) the alignment using only rigid transformations and respectively in Figure 2(b) the result after applying the methodology presented in this paper can be seen. Unfortunately we can only demonstrate the output of our method visually, because required ground truth model was not obtained in the same conditions as the patient was scanned.



(a) (b)

Figure 2: Non-rigid registration of patient arm, red arrow shows the gap between two rigid registered point clouds; a) rigid registration; b) registration with proposed method.

5 Conclusion and Future Work

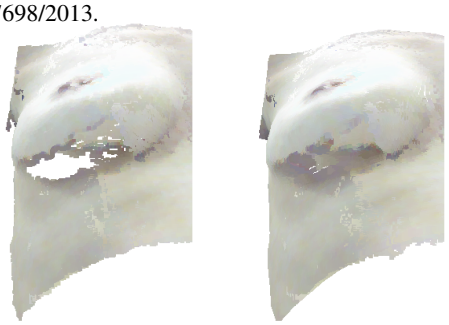
Although rigid registration methodologies are not capable to present an accurate registration in the presence of non-linear differences, by utilizing NURBS technique it is possible to assure the alignment of point clouds. In this paper we proposed a simple method to manipulate point clouds individually in order to align them, even in the presence of non-linear misalignments. Experimental results indicate that proposed method aligns the views, in term of decreasing residual distance between registered point clouds; however, quantitative evaluation using ground truth data is necessary to properly validate the implemented methodology.

Additionally, and using only the regularization step of the method, the proposed method can be used to fill missing areas of point cloud, since NURBS generates uniform surfaces. In Figure 3 we can visualize the result after applying the proposed approach, to point cloud breast data acquired with Kinect device.

Future work will be focused on two aspects: conducting more accurate non-rigid registration, and improving the performance of hole filling. Non-rigid registration can be enhanced by using color information of the points in order to select correct corresponding patches. Besides, value used to define maximum allowed distance between patches can be obtained using point cloud density, rather than using predefined values. On the other hand, inserting control points in correct positions guarantee the accurate filling of missing points in order to preserve deformation of reconstructed model.

Acknowledgment

This work is financed by the European Community's Seventh Framework Programme [grant number FP7600948], by the ERDF European Regional Development Fund through the COMPETE Programme (operational programme for competitiveness), FCOMP-01-0124-FEDER-037281 and by National Funds through the Fundação para a Ciência e a Tecnologia within project PEst-C/EEI/LA0014/2013 and the PhD grant number SFRH/BD/97698/2013.



(a) (b)

Figure 3: Filling holes using interpolation of the NURBS generator; a) point cloud wit holes in the bottom of breast; b) point cloud after applying proposed method. Note that breast beneath is reconstructed with darker color because it was shadowed by breast during scanning procedure.

References

[1] Pedro Costa, João P. Monteiro, Hooshiar Zolfagharnasab, and Hélder P. Oliveira. Tessellation-based coarse registration method for 3d reconstruction of the female torso. In *Proceedings of the IEEE International Conference on Bioinformatics and Biomedicine*. 2014.

[2] Pedro Costa, Hooshiar Zolfagharnasab, João P. Monteiro, Jaime S. Cardoso, and Hélder P. Oliveira. 3d reconstruction of body parts using rgb-d sensors: Challenges from a biomedical perspective. In *Proceedings of the 5th International Conference on 3D Body Scanning Technologies*. 2014.

[3] Nallig Leal, Esmeide Leal, and John William Branch. Simple method for constructing nurbs surfaces from unorganized points. In *Proceedings of the 19th International Meshing Roundtable*, pages 161–175. 2010.

[4] Hao Li, Etienne Vouga, Anton Gudym, Linjie Luo, Jonathan T. Barron, and Gleb Gusev. 3d self-portraits. *ACM Trans. Graph.*, 32(6): 187:1–187:9, 2013.

[5] Hélder P. Oliveira, Jaime S. Cardoso, André T. Magalhaes, and Maria J. Cardoso. A 3d low-cost solution for the aesthetic evaluation of breast cancer conservative treatment. *Computer Methods in Biomechanics and Biomedical Engineering: Imaging and Visualization*, 2(2):90–106, 2014.

[6] Jianzhe Wang and Tianzi Jiang. Nonrigid registration of brain mri using nurbs. *Pattern Recogn. Lett.*, 28(2):214–223, 2007.

Multi-Target Tracking Based in Meanshift and Particle Filters

Pedro Silva, Ricardo Ferreira, José Gaspar
http://www.isr.ist.utl.pt

Institute for Systems and Robotics / IST
Lisbon, Portugal

Abstract

In this work we propose a hybrid particle filter and meanshift tracker to follow multiple targets using one camera. The number of targets is assumed to be known but allowed to vary in time.

1 Introduction

The kernel based-tracking algorithm was first proposed in [1], and despite its simplicity it proved to give good results even in the presence of motion blur and clutter. However it had the drawback of assuming that the target did not move outside of the region of interest between consecutive frames. Years later, Maggio et al. [2] proposes an hybrid target tracking algorithm that by using a particle filter and the kernel based tracking algorithm was able to increase the robustness to the mentioned problem. Still it was designed for single target tracking. The algorithm proposed in this paper, therefore intends to improve the work of [2] by addressing the multi-target tracking problem.

2 Camera Model

We use the pinhole camera model [3] to derive the perspective transformation of the points on a world referential frame to the ones in the camera image plane,

$$m \approx PM = K \begin{bmatrix} {}^{C}R_W \mid {}^{C}t_W \end{bmatrix} M, \tag{1}$$

where M is a vector containing the coordinates of the point in the world referential frame, m are the pixel coordinates, P is the projection matrix, K and ${}^{C}R_W$ are the intrinsic and extrinsic parameters matrices respectively and finally ${}^{C}t_W$ is the translation vector. Given the projection matrix and a 3D plane an image pixel m can also be back-projected to a point $M = [x\ y\ z\ 1]^\intercal$ in the plane.

3 Target Description

The state of each target to track is described, at the time step t, by the vector $\mathbf{x}_t = (x_t, y_t, \theta_t, v_t)$, where x_t, y_t are the target spatial coordinates in the world referential frame and θ_t, v_t the target orientation and linear velocity respectively. The targets are assumed to be moving in a plane, where their z coordinate is constant and computed at the moment they are detected. Furthermore the number of targets being tracked, at a given time step, is assumed to be known.

In order to make predictions of a target next state the kinematic model $\mathbf{x}_t = f(\mathbf{x}_{t-1}) + \mathbf{w}_t$ is used with

$$f(\mathbf{x}_t) = \begin{cases} x_t = x_{t-1} + T v_{t-1} \cos(\theta_{t-1}) \\ y_t = y_{t-1} + T v_{t-1} \sin(\theta_{t-1}) \\ \theta_{t-1} = \theta_{t-1} \\ v_t = v_{t-1} \end{cases}, \tag{2}$$

T is the time step value and $\mathbf{w}_t \sim \mathcal{N}(\mathbf{0}, Q)$ an additive noise component drawn from the multivariate Gaussian distribution $\mathcal{N}$, with mean equal to the null vector $\mathbf{0}$ and covariance matrix Q. The function $f(\mathbf{x}_t)$ is derived from the car model used in [4], by setting to zero the wheel orientation angle ϕ. In addition to the kinematic model a target has an appearance model, under the form of a RGB color histogram, which allows to distinguish the target in the image data. This model is used to produce estimations of the targets position and orientation, by searching the image data. It is computed and assigned when the target is first detected.

4 Optimal Proposal Particle Filter

The particle filter is a Monte Carlo [5] based tracking filter used to recursively estimate the posterior probability density function $p(\mathbf{x}_t|\mathbf{z}_{1:t})$, where $\mathbf{z}_{1:t}$ are the set of observation vectors $\mathbf{z}$ up to time t. In the case of this tracking filter the measurement and kinematic models do not need to be linear, and the noise can be non additive and non-Gaussian.

The particle filter algorithm starts by drawing a set of N samples $\mathbf{x}_t^i, i = 1,...,N$ from an importance distribution $\mathbf{x}_t^i \sim q(\mathbf{x}_t^i|\mathbf{x}_{t-1}^i, \mathbf{z}_t)$. To each of these samples a weight ω_t^i is assigned through the formula:

$$\omega_t^i = C\omega_{t-1}^i \frac{p(\mathbf{z}_t|\mathbf{x}_t^i)p(\mathbf{x}_t^i|\mathbf{x}_{t-1}^i)}{q(\mathbf{x}_t^i|\mathbf{x}_{t-1}^i, \mathbf{z}_t)}, \tag{3}$$

where $p(\mathbf{x}_t^i|\mathbf{x}_{t-1}^i)$ denotes the prior distribution, $p(\mathbf{z}_t|\mathbf{x}_t^i)$ is the likelihood and C a normalization constant to make the sum of the weights equal to 1. These weights can then be used to make the following discrete approximation of the posterior distribution:

$$p(\mathbf{x}_t|\mathbf{z}_{1:t}) \approx \sum_{i=1}^{N} \omega_t^i \delta(\mathbf{x}_t - \mathbf{x}_t^i) \tag{4}$$

One commonly used estimator of the state at time t is

$$\mathbb{E}(p(\mathbf{x}_t|\mathbf{z}_{1:t})) \approx \frac{1}{N} \sum_{i=1}^{N} \omega_t^i \mathbf{x}_t^i, \tag{5}$$

where $\mathbb{E}$ denotes the expectation operator. Resampling, an additional step usually included in particle filters, replaces the actual samples by a set of new ones drawn from the discrete approximation of the posterior, hence propagating to the next iteration a set of samples that more likely represent the true system state.

In one of the simplest particle filter implementations, commonly called the bootstrap particle filter, the importance distribution is chosen to be the prior distribution: $q(\mathbf{x}_t^i|\mathbf{x}_{t-1}^i, \mathbf{z}_t) = p(\mathbf{x}_t^i|\mathbf{x}_{t-1}^i)$. This simplifies expression (3) making $\omega_t^i = Cp(\mathbf{z}_t|\mathbf{x}_t^i)$. However a common issue with this approach is a phenomenon typically known as degeneracy, that is characterized by the tendency for a few set of particles to have weights much larger than the rest. This leads to a poor approximation of the posterior distribution. One possible solution to overcome this is to choose a better importance sampling. It has been proven [5] that the choice $q(\mathbf{x}_t^i|\mathbf{x}_{t-1}^i, \mathbf{z}_t) = p(\mathbf{x}_t^i|\mathbf{x}_{t-1}^i, \mathbf{z}_t)$ is optimal in the sense of minimizing the variance of the particles weights. The great disadvantage of this choice is that it is usually very difficult to evaluate. However for systems with models of the form

$$\begin{matrix} \mathbf{x}_t = f(\mathbf{x}_{t-1}) + \mathbf{w}_t,\ \mathbf{w}_t \sim \mathcal{N}(\mathbf{0}, Q) \\ \mathbf{z}_t = H\mathbf{x}_t + \nu_t,\ \nu_t \sim \mathcal{N}(\mathbf{0}, R) \end{matrix}, \tag{6}$$

with R being the measurement covariance matrix and H the observation matrix, it is possible to derive closed form expressions for $p(\mathbf{z}_t|\mathbf{x}_{t-1})$ and $p(\mathbf{x}_t|\mathbf{x}_{t-1}, \mathbf{z}_t)$ which, as deduced in [5], are:

$$p(\mathbf{z}_t|\mathbf{x}_{t-1}) = \mathcal{N}(\mathbf{b}_t, S) \tag{7}$$

$$p(\mathbf{x}_t|\mathbf{x}_{t-1}, \mathbf{z}_t) = \mathcal{N}(\mathbf{a}_t, \Sigma) \tag{8}$$

$$\mathbf{b}_t = Hf(\mathbf{x}_{t-1}) \tag{9}$$

$$S = HQH^\intercal + R \tag{10}$$

$$\mathbf{a} = f(\mathbf{x}_{t-1}) + \Sigma H^\intercal R^{-1}(\mathbf{z}_t - b_t) \tag{11}$$

$$\Sigma = Q - QH^\intercal S^{-1} HQ \tag{12}$$

5 Multi-target Hybrid Particle Filter

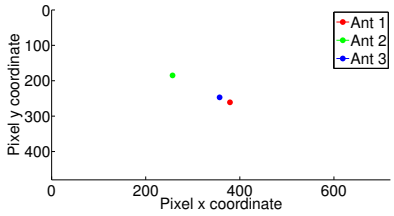
The first step in the tracking algorithm proposed is similar to the standard *prior* update of a bootstrap particle filter where the kinematic model is used to update the particles state. The difference lies in the fact that the noise is not added, with only the function $f(\cdot)$ being employed.

Then a procedure inspired by the work of Maggio et al. [2] is applied to each target, in order to get observations for its position. It starts by selecting a smaller number of particles, picked at random from the target particle set with the probability of each being drawn equal to its respective weight. If there are any duplicates they are removed in order to only have

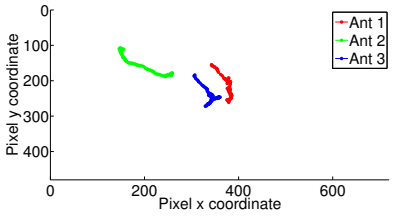
(a) Frame 1

(b) Frame 97

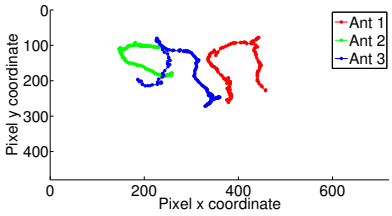
(c) Frame 197



(d) Frame 1



(e) Tracking till frame 97



(f) Tracking till frame 197

Figure 1: Tracking twenty ants. Displaying just three for clarity.

unique particle states. The full set of particles is not used only to decrease the computational burden.

The x and y hypothesis contained in each particle, together with the constant z coordinate of the target, are then used as the center of an ellipse, whose rotation around the z axis is set equal to the target orientation estimated in the previous iteration. The ellipse is then used for the application of the kernel based tracking method of Comaniciu et al. [1], together with the target appearance model. This is a method that, using the mean-shift algorithm, iteratively moves the elliptical region to a location in the image, where the RGB histogram of the pixels inside it more closely resembles the target RGB histogram. The similarity measure used is the Bhattacharyya coefficient. Finally the convergence points obtained are back-projected into the world referential frame and stored in the set $\mathcal{O}_t$, together with those of the targets that have already been iterated.

The points contained in $\mathcal{O}_t$ are then clustered using a mixture of Gaussians, with the number of clusters being equal to the number of targets, N_T. The mixture of Gaussians means are then stored to be used as observations for the posterior update of the optimal proposal particle filter. It is however necessary to associate these observations to their respective target. For that a Global Nearest Neighbours (GNN) [6] approach is employed by using the Hungarian algorithm [7], where each element of the cost matrix is equal to the distance between the observation, associated to each row, and the previous estimation of a targets position, associated to the columns. This process of clustering intends to add robustness to the problem that occurs when two similar targets are close to each other and the elliptical region of one target is attracted to the other one since this happens to have a higher Bhattacharyya coefficient.

With the observations associated to their respective targets a last step is done to obtain observations for the targets orientation. This is done through an exhaustive search in which an ellipse centered at each of the observations is placed in different rotation values. The one which gives a greater Bhattacharyya coefficient is used as the orientation for that particular observation. With the observations finally obtained, the likelihood is computed for each particle through the measurement equation $\mathbf{z}_t = \begin{bmatrix} I \mid & \mathbf{0} \end{bmatrix} \mathbf{x}_t + v_t, v_t \sim \mathcal{N}(\mathbf{0}, R)$, where I is the identity matrix.

6 Experimental Results

The proposed algorithm was applied to a video were twenty ants are moving in a dish [8]. This video has a dimension of 720×480 pixels and a duration of about eight seconds. The initial positions of the ants are set manually and sent to the algorithm.

In order to analyze the quality of the results a comparison is made with a multi-target tracker using background subtraction to detect the ants and the Multi-hypothesis tracker (MHT) algorithm for tracking and data association [9]. However as there are a considerable number of ants, only the trajectories and error values of three randomly selected ants are presented. The ground truth was manually extracted from the video.

In figures 1(a) to 1(c) three frames of the ants video are shown. Associated to them the trajectories estimated by the method proposed, named Multi-target Hybrid Particle Filter (MHPF), to the respective frame. The table 1 shows the root mean square error (RMSE), the total number of tracks lost (TL) and the number of tracks recovered (TR), for the several ants and the two algorithms.

	MHT			MHPF		
	Ant1	Ant2	Ant3	Ant1	Ant2	Ant3
RMSE	6.35	2.54	4.39	4.14	1.95	2.88
TL	152			0		
TR	152			-		

Table 1: Performance measures between the two algorithms

A track is considered lost when it changes the label that it uses to identify a target. As can be seen the MHPF showed better results in comparison with the other algorithm. Not only the RMSE is lower for all the ants but it also did not lose any track.

7 Conclusion and Future Work

The proposed tracking algorithm proved to give good results, even when the targets being tracked have all the same appearance. However it still needs further improvement, especially in the handling of total occlusions.

Acknowledgments

This work has been partially supported by the FCT project PEst-OE / EEI / LA0009 / 2013, by the QREN project 34063 SEAGULL, and by the EU project POETICON++ FP7-ICT-288382.

References

[1] Dorin Comaniciu, Visvanathan Ramesh, and Peter Meer. Kernel-based object tracking. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 25(5):564–577, 2003.

[2] Emilio Maggio and Andrea Cavallaro. Hybrid particle filter and mean shift tracker with adaptive transition model. In *ICASSP (2)*, pages 221–224, 2005.

[3] R. Hartley and A. Zisserman. *Multiple View Geometry in Computer Vision*. Cambridge University Press, 2000.

[4] Tiago Castanheira. Multitasking of smart cameras. *Dissertação para a obtenção do Grau de Mestre em Engenharia Electrotécnica e de Computadores*, 2013.

[5] Anton J Haug. *Bayesian Estimation and Tracking: A Practical Guide*. John Wiley & Sons, 2012.

[6] A. Udvarev P. Konstantinova and T. Semerdjiev. A study of a target tracking algorithm using global nearest neighbor approach. In *Proceedings of the International Conference on Computer Systems and Technologies (CompSysTech)*, 2003.

[7] Harold W Kuhn. The hungarian method for the assignment problem. *Naval research logistics quarterly*, 2(1-2):83–97, 1955.

[8] Zia Khan, Tucker Balch, and Frank Dellaert. An MCMC-based particle filter for tracking multiple interacting targets. In *Computer Vision-ECCV 2004*, pages 279–290. Springer, 2004.

[9] David Miguel Antunes, David Martins de Matos, and José Gaspar. A library for implementing the multiple hypothesis tracking algorithm. *arXiv preprint arXiv:1106.2263*, 2011.

Large-scale inversions between human reference assemblies

Diogo Pratas
pratas@ua.pt
Raquel M. Silva
raquelsilva@ua.pt
Armando J. Pinho
ap@ua.pt

Signal Processing Lab, IEETA / DETI,
University of Aveiro,
3810–193 Aveiro, Portugal

Abstract

The detection of large-scale inversions between intra-species genomes is fundamental to understand the dynamics of chromosomal evolution, namely to identify hallmarks associated with diseases. In this paper, we explore a method that finds exact or approximate large-scale inversions, for regions larger than 500 kb, between the three most popular assembled human genomes. We identify the relative positions for each chromosome. The inversions are computed in two maps showing at a glance the associated regions.

Introduction

How genome architecture and which macroevolutionary events emerge through time are fundamental to understand the dynamics of species evolution, namely their origin and speciation patterns [3]. Several insights into chromosome evolution have been traditionally achieved by cytogenetic procedures, for example G-banding, and by molecular karyotyping approaches, such as fluorescence in situ hybridization (FISH). More recently, array-based methods became very popular [2].

Advances in sequencing technology have increased the number of digital human genomes, raising conditions towards intra-species characteristics and diversity research. Consequently, computational approaches emerged [5], bringing high resolution, accuracy and speed with less costs. However, the de novo assembly of the next generation sequencing (NGS) reads is still problematic, mainly because the alignment of the reads from these new genomes to a high quality reference genome remains a critical aspect of data interpretation. Nevertheless, the human reference assembly is the highest quality mammalian assembly available. The main reference genome assemblies are those from the Genome Reference Consortium (GRC 38) [1], the J. Craig Venter Institute (HuRef) [4] and the Washington U. School of Medicine (CHM 1.1).

In this paper, we detect the large-scale (larger than 500 Kb) inversions between these three reference genome assemblies, exploring fundaments from an unsupervised alignment-free method [9], that is based in the hability to efficiently model the repetitiveness of genomic sequences [8].

Method

The method involves the estimation of the amount of conditional exclusive information, i. e., using only information from a reference that is required to represent a certain region of the sequence or sub-sequence [6]. The process is as follows:

1. Convert symbols outside $\mathcal{A} = \{A, C, G, T, N\}$ into "N";
2. Pseudo-randomize the "N" symbols (with uniform distribution);
3. Invert the first sequence;
4. Load to the FCM [7] the sequence created in step 3;
5. Compress the second sequence using the FCM created in step 4;
6. Filter the information sequence generated in 5;
7. Use a threshold of 1.6 to segment the sub-sequence regions;
8. Paint sub-sequence regions with different colors;
9. For each sub-sequence region, repeat the process from step 3 to 8, using the second sequence as target;

The colors are calculated automatically using a HSV scheme where only V varies.

Results

For the results presented in this section we have downloaded the genomes from the NCBI[1] site. Chromosome Y from CHM genome is absent and thus we have not reported results associated with this sequence.

Fig. 1 shows the maps from the large-scale inversions between A (GRC) and B (HuRef), while Fig. 2 shows the inversions between A(GRC) and C (CHM) assemblies. In respect to A/B, there are inversions in chromosomes 1, 2, 7, 9, 10, 11, 15 and Y. Specifically to chromosome 1, the inversions are contained in the pericentric regions (around 119 Mb). This region is also inverted between human and chimpanzee species, although in a much larger density [9].

On the other hand, for A/C the inversions are present between chromosomes 1, 2, 5, 7, 8, 9, 10, 11, 14, 15, 16, 17, 22, X. From these, most of the inversions are contained in pericentric regions, a major factor of dynamism across individuals of the same species.

Conclusions

We have proposed a procedure to detect exact or approximate inversions, larger than 500 Kb, between reference assembled genomes. The inversions have been computed in a information map.

Besides specific characteristics, the reference assembly from the Washington U. School of Medicine (CHM) seems to have a higher number of inversions compared with the Genome Reference Consortium (GRC 38), than with the J. Craig Venter Institute (HuRef).

Acknowledgements

Supported by the European Fund for Regional Development (FEDER) through the Operational Program Competitiveness Factors (COMPETE) and by the Portuguese Foundation for Science and Technology (FCT), in the context of projects PEst-OE/EEI/UI0127/2014 and Incentivo/EEI/UI0127/2014. DP is supported by the European Union Seventh Framework Programme (FP7/2007-2013) under grant agreement No. 305444 "RD-Connect: An integrated platform connecting registries, biobanks and clinical bioinformatics for rare disease research". RMS is supported by the project Neuropath (CENTRO-07-ST24-FEDER-002034), co-funded by QREN "Mais Centro" program and the EU.

References

[1] D. Church, M. Deanna, V. Schneider, et al. Modernizing reference genome assemblies. *PLoS Biology*, 9(7):e1001091, 2011.

[2] K. Das and P. Tan. Molecular cytogenetics: recent developments and applications in cancer. *Clinical genetics*, 84(4):315–325, 2013.

[3] Marta Farré and Aurora Ruiz-Herrera. Role of chromosomal reorganisations in the human–chimpanzee speciation. *eLS*, 2014.

[4] S. Levy, G. Sutton, P. C. Ng, L. Feuk, A. L. Halpern, et al. The diploid genome sequence of an individual human. *PLoS Biology*, 5:2113–2144, 2007.

[5] C. Nielsen, M. Cantor, I. Dubchak, D. Gordon, and T. Wang. Visualizing genomes: techniques and challenges. *Nature methods*, 7:S5–S15, 2010.

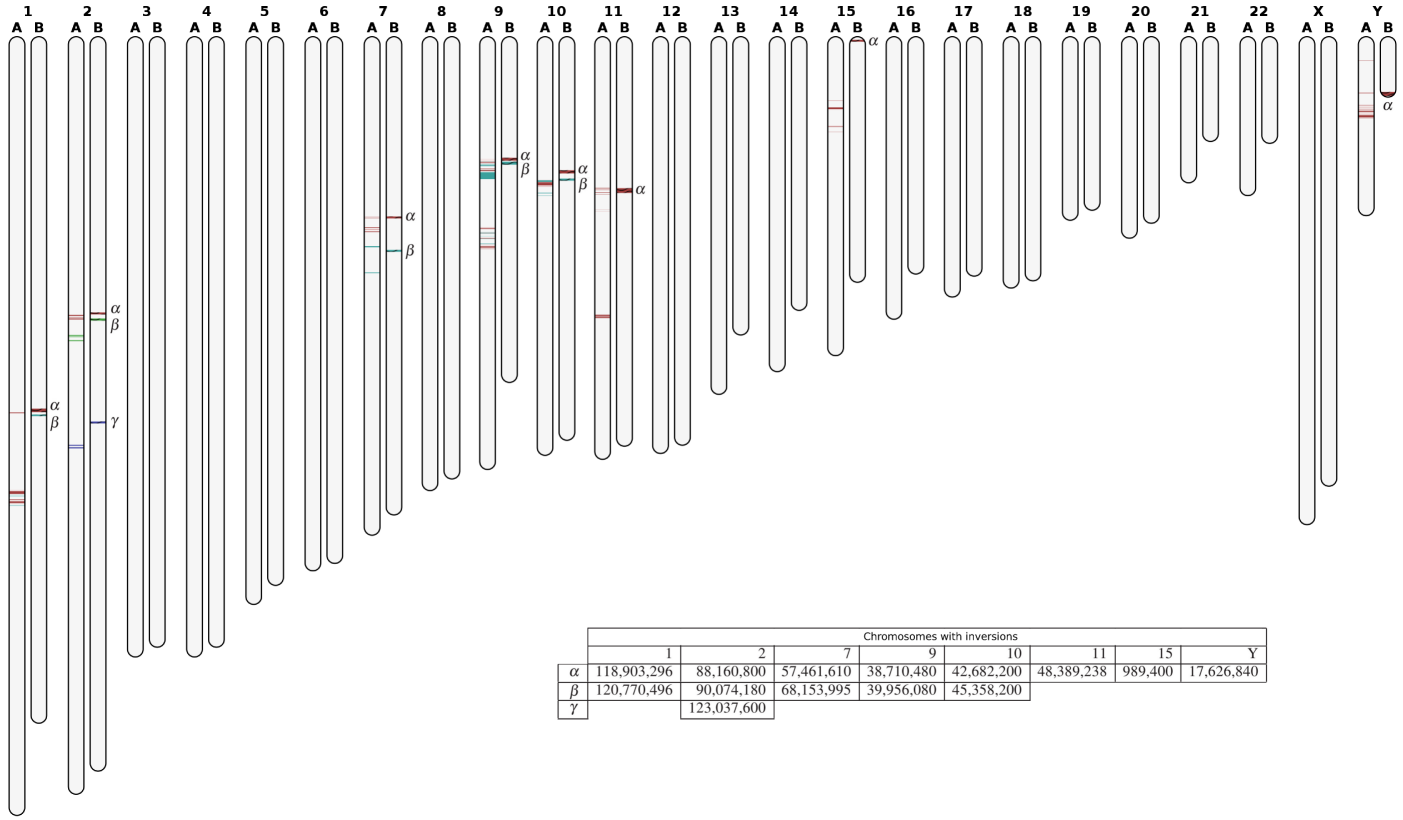
[6] A. J. Pinho, D. Pratas, P. J. S. G. Ferreira, and S. P. Garcia. Symbolic to numerical conversion of DNA sequences using finite-context models. In *Proc. of the 19th European Signal Processing Conf., EUSIPCO-2011*, Barcelona, Spain, August 2011.

[7] A. J. Pinho, D. Pratas, and S. P. Garcia. Complexity profiles of DNA sequences using finite-context models. In *Information Quality in e-Health*, volume 7058, pages 75–82. Springer, 2011.

[8] A. J. Pinho, S. P. Garcia, D. Pratas, and P. J. S. G. Ferreira. DNA sequences at a glance. *PLoS ONE*, 8(11):e79922, November 2013.

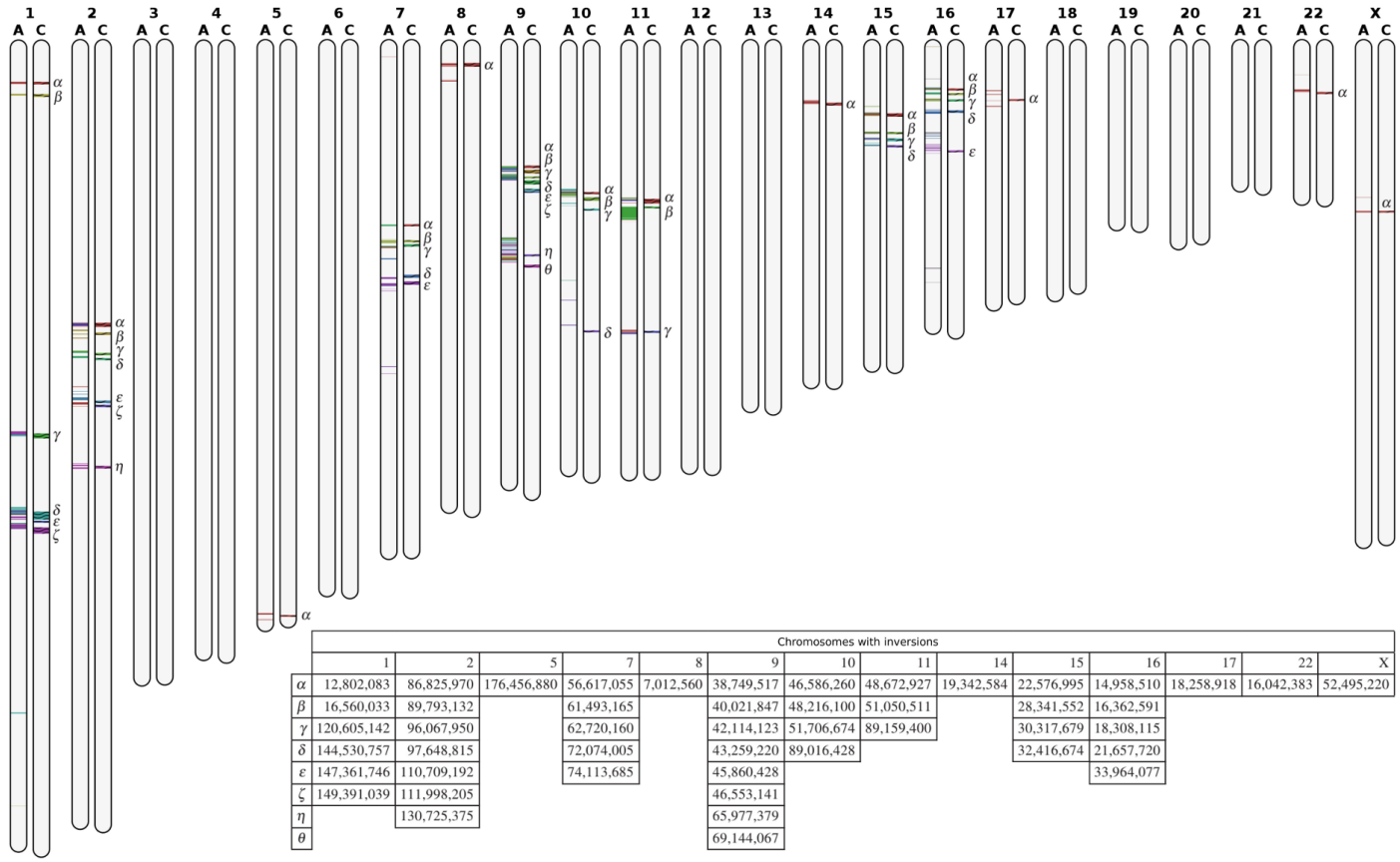
[9] D. Pratas, R. M. Silva, A. J. Pinho, and P. J. S. G. Ferreira. An alignment-free method to find and visualize genomic rearrangements. *Submitted*, 2014.

[1] `ftp://ftp.ncbi.nlm.nih.gov/genomes/Homo_sapiens/Assembled_chromosomes/seq/`



	Chromosomes with inversions							
	1	2	7	9	10	11	15	Y
α	118,903,296	88,160,800	57,461,610	38,710,480	42,682,200	48,389,238	989,400	17,626,840
β	120,770,496	90,074,180	68,153,995	39,956,080	45,358,200			
γ		123,037,600						

Figure 1: Large-scale inversions between GRC (A) and HuRef (B) assemblies for each chromosome. The information maps show exact or approximate inversions with length higher than 500 kb. Each position associated with inversions, in the HuRef chromosomes, is reported in the table and marked with a greek letter according to the map.



	Chromosomes with inversions													
	1	2	5	7	8	9	10	11	14	15	16	17	22	X
α	12,802,083	86,825,970	176,456,880	56,617,055	7,012,560	38,749,517	46,586,260	48,672,927	19,342,584	22,576,995	14,958,510	18,258,918	16,042,383	52,495,220
β	16,560,033	89,793,132		61,493,165		40,021,847	48,216,100	51,050,511		28,341,552	16,362,591			
γ	120,605,142	96,067,950		62,720,160		42,114,123	51,706,674	89,159,400		30,317,679	18,308,115			
δ	144,530,757	97,648,815		72,074,005		43,259,220	89,016,428			32,416,674	21,657,720			
ε	147,361,746	110,709,192		74,113,685		45,860,428					33,964,077			
ζ	149,391,039	111,998,205				46,553,141								
η		130,725,375				65,977,379								
θ						69,144,067								

Figure 2: Large-scale inversions between GRC (A) and CHM (C) assemblies for each chromosome. The information maps show exact or approximate inversions with length higher than 500 kb. Each position associated with inversions, in the CHM chromosomes, is reported in the table and marked with a greek letter according to the map.

Towards personalized medicine: ebola virus absent words in the human genome

Raquel M. Silva
raquelsilva@ua.pt
Luisa Castro
luisa.castro@ua.pt
Diogo Pratas
pratas@ua.pt
Armando J. Pinho
ap@ua.pt

Signal Processing Lab, IEETA / DETI,
University of Aveiro,
3810–193 Aveiro, Portugal

Abstract

Next generation sequencing technologies are driving a novel revolution in the personalized medicine field. The unprecedented availability of genomes and computational tools enables the development of novel diagnosis and therapeutic strategies. Here, using an alignment-free method based on the relative minimal absent words, we show that the identification of pathogen signatures is possible to allow quick intervention for infectious agents, as exemplified by the current Ebola virus outbreak genome analysis.

Introduction

The $1200 genome milestone has been reached, with sequencing times of one week, and it has been argued that these values will substantially decrease with the incoming of the third generation. The personalized medicine field is now newborn, increasingly dependent on the advances of sequencing technologies, and yielding major contributions for diagnosis or genetic counseling to maximize the probability of welfare. Under this framework, personalized vaccines are a possibility and their development is essential with the emergence of pathogen resistance, namely for fungal, viral and bacterial infections.

New successful applications strongly depend on the ability to detect regions in pathogen genomes that are absent in the healthy host, mainly to eliminate the pathogen without harming the host. Given the task, genomic information sizes create a challenge to the wet lab, and currently this can only be solved with comparative genomics that can be time-consuming, expensive and more susceptible to error than computational approaches.

Although computational methods can be employed to solve these tasks, the success of the detection algorithms are strongly dependent on the time and memory requirements to run on common computers. Several studies indicate that minimal absent words on the human genome exist above size 10. Therefore, efficient data structures, such as suffix arrays and hash-tables, among others, play a key role to minimize memory requirements maintaining affordable processing times.

On the other hand, identifying comparative specific novel regions is also a way to discover new genes and genome structures, and evidence of evolutionary patterns and signatures across species. For example, one of the current research topics is the identification of modern human specific genes, compared with the newly sequenced genomes of Denisovans and Neanderthals. Typically, these genomes have approximately 3 giga bases, establishing the importance of efficient fast-compact data structures.

Minimal absent words have been studied and computed by many researchers [1, 3, 5, 6]. In this paper, we follow an emerging branch, the relative minimal absent words. These are minimal absent words of a certain genome that exist in another one. We propose a method to detect these words and report their position. We have applied the method to 20 Ebola virus (EBOV) genomes and detected several minimal absent words of the human reference genome (GRC) that occur in the virus, as well as their positions.

Method

Consider a reference sequence, X, and n target sequences $Y_1, Y_2, \ldots, Y_n$. All sequences are from a finite alphabet, $\Sigma = \{A, C, G, T\}$, and $|X|$ denotes the size of sequence X.

We compute the k-mers of X and the Y_i sequences using a sliding window of size k. Each k-mer is converted into a numeric index, i, and stored into a binary array if $k < 17$, otherwise in a hash table. Parallel, we perform the following mapping: $A \rightarrow T$, $T \rightarrow A$, $C \rightarrow G$ and $G \rightarrow C$. The mapping is applied for each reversed k-mer, converted into an index, i and stored as described above. Each k-mer from X is loaded, including those from the reverse mapping, and stored in memory. We call this the training phase.

Then we start the matching phase. The intention is to find exact k-mers on each Y_i. Therefore, for each Y_i, a boolean array is created, B_i, with $|B_i| = |Y_i| - k$, containing a true value when a k-mer exists in the memory.

The objective is to detect relative absent words, therefore the interest is on the false elements from B_i. Since the process of matching is sequential, each position of a false element in B_i reports the exact position in the target sequence, Y_i.

Finally, the results are presented in a map, along each Y_i, that depicts the regions or points where the k-mer (absent in X) occurs.

Results

For the results in this section we have used the full GRC-38 human reference genome [2] downloaded from the NCBI[1], including the mitochondrial, unplaced and unlocalized sequences. For the 20 EBOV genomes, the sequences (first 20 genomes) have been also downloaded from NCBI[2] [4].

We have implemented a version of the method that required 7 minutes and 51 seconds to compute all the k-mers presented in Fig. 1, from the 20 EBOV genomes, including the training phase over a sequence of approximately 3 GB. The maximum RAM memory used was 1 GB. The computation was performed on a Linux Ubuntu 12.04 LTS with the following hardware characteristics: 4 Intel Core i7-3520M CPU at 2.90GHz, 8 GB of RAM and a SSD of 243.6 GB.

Figure 1 depicts the computation for word sizes 12, 13, 14 and 15. As expected, the number of absent words decreases as the k-mer size decreases. Specifically, for $k = 11$ (not in Fig. 1), there are no relative absent words. On the other hand, for $k = 12$, three groups of points emerge (P1, P2 and P3). Each position of the corresponding point, according to each genome, is shown in Table 1. In each group, the positions are very close among the different strains of the virus.

In fact, they all degenerate from the same words, namely:

- TTTCGCCCGACT (P1),
- TACGCCCTATCG (P2),
- CCTACGCGCAAA (P3).

From the three EBOV sequence motifs absent in the human genome, the first (P1) is included in the virus nucleoprotein (NP), while the other two (P2 and P3) fall within the sequence of the viral RNA-polymerase (L protein). Previous studies show that the N-terminal region of EBOV NP participates in both the formation of nucleocapsid-like structures through NP-NP interactions and in the replication of the viral genome [7], and P1 sequence is part of this N-terminal region (Fig. 2). The L-protein produces the viral transcripts to be translated by host ribosomes and is involved in the replication of the viral genome as well. Both proteins are critical for the virus life cycle, thus, constitute good targets for therapeutic intervention. The identification of these viral genome signatures is also important for quick diagnosis in outbreak scenarios.

[1] ftp://ftp.ncbi.nlm.nih.gov/genomes/H_sapiens/Assembled_chromosomes/seq/

[2] http://www.ncbi.nlm.nih.gov/bioproject/257197

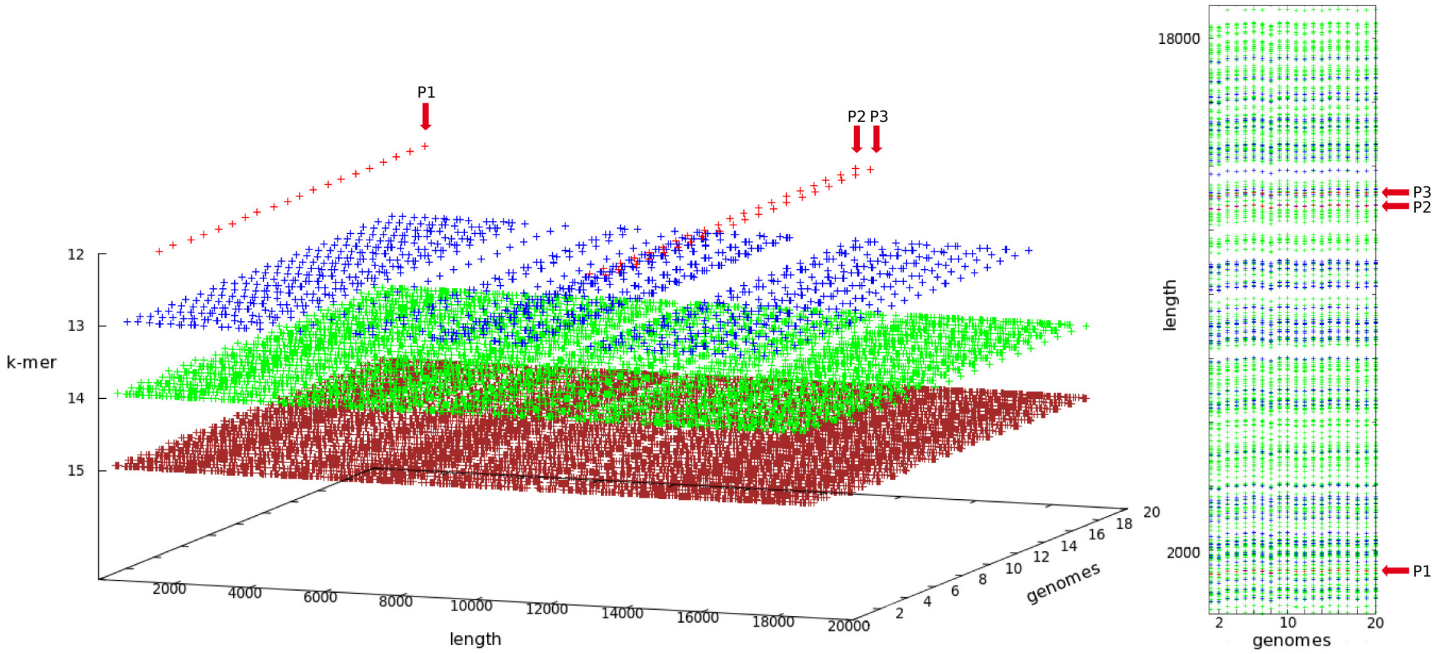


Figure 1: Ebola virus absent words relatively to the GRC reference human genome. Left plot depicts a 3D model with several k-mers for the 20 EBOV genomes along their length, while the right plot contains a (vertical) projection of the genomes and length. Points $P1$, $P2$ and $P3$ represent the relative minimal absent words.

Figure 2: Structure of the Ebola virus genome. The negative-stranded RNA genome has about 19 kb in size and encodes for seven proteins: nucleoprotein (N), glycoprotein (GP), polymerase (L) and four additional viral proteins (VP).

Genome	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
P1	1,232	1,216	1,323	1,308	1,336	1,353	1,318	1,260	1,351	1,353	1,309	1,308	1,340	1,309	1,353	1,345	1,340	1,308	1,329	1,347
P2	12,654	12,638	12,745	12,730	12,758	12,775	12,740	12,682	12,773	12,775	12,731	12,730	12,762	12,731	12,775	12,767	12,762	12,730	12,751	12,769
P3	13,056	13,040	13,147	13,132	13,160	13,177	13,142	13,084	13,175	13,177	13,133	13,132	13,164	13,133	13,177	13,169	13,164	13,132	13,153	13,171

Table 1: Starting positions for each absent word, for $k = 12$, contained in each EBOV genome relatively to the GRC reference human genome.

Conclusions

The field of personalized diagnosis and terapeutics is largely led by the virology branch. The identification of the regions that are present in a virus genome but are absent in a human genome, will drive the development of innovative terapeutics.

It is well-known that two different human genomes share a high degree of homology, and we explore this characteristic to detect minimal absent words in the human genome that are present in the Ebola virus. For each of the 20 EBOV genomes, we have detected 3 words with size 12, namely, *TTTCGCCCGACT*, *TACGCCCTATCG* and *CCTACGCGCAAA*, using the method that we propose here.

These results can now be further explored from a biological point of view, in order to build a vaccine that has a high probability to damage the virus without harming the human being.

Acknowledgements

Supported by the European Fund for Regional Development (FEDER) through the Operational Program Competitiveness Factors (COMPETE) and by the Portuguese Foundation for Science and Technology (FCT), in the context of projects PEst-OE/EEI/UI0127/2014 and Incentivo/EEI/-UI0127/2014. RMS and LC are supported by the project Neuropath (ref: CENTRO-07-ST24-FEDER-002034), co-funded by QREN "Mais Centro" program and the EU. DP is supported by the European Union Seventh Framework Programme (FP7/2007-2013) under grant agreement No. 305444. Info:"RD-Connect: An integrated platform connecting registries, biobanks and clinical bioinformatics for rare disease research".

References

[1] Supaporn Chairungsee and Maxime Crochemore. Using minimal absent words to build phylogeny. *Theoretical Computer Science*, 450: 109–116, 2012.

[2] D. Church, M. Deanna, V. Schneider, et al. Modernizing reference genome assemblies. *PLoS Biology*, 9(7):e1001091, 2011.

[3] S. P. Garcia, A. J. Pinho, J. M. O. S. Rodrigues, C. A. C. Bastos, and P. J. S. G. Ferreira. Minimal absent words in prokaryotic and eukaryotic genomes. *PLoS ONE*, 6(1):e16065, January 2011. doi: doi:10.1371/journal.pone.0016065.

[4] Stephen K Gire, Augustine Goba, Kristian G Andersen, Rachel SG Sealfon, Daniel J Park, Lansana Kanneh, Simbirie Jalloh, Mambu Momoh, Mohamed Fullah, Gytis Dudas, et al. Genomic surveillance elucidates ebola virus origin and transmission during the 2014 outbreak. *Science*, page 1259657, 2014.

[5] J. Herold, S. Kurtz, and R. Giegerich. Efficient computation of absent words in genomic sequences. *BMC Bioinformatics*, 9(1):167, 2008. doi: 10.1186/1471-2105-9-167.

[6] A. J. Pinho, P. J. S. G. Ferreira, S. P. Garcia, and J. M. O. S. Rodrigues. On finding minimal absent words. *BMC Bioinformatics*, 10 (137), May 2009. doi: 10.1186/1471-2105-10-137.

[7] Shinji Watanabe, Takeshi Noda, and Yoshihiro Kawaoka. Functional mapping of the nucleoprotein of ebola virus. *Journal of virology*, 80 (8):3743–3751, 2006.

A Cellular Automaton for Enzyme Kinetics

João Silva[1]

Nelson Fernandes[3]

Paulo Fazendeiro[1]

António Mendonça[23]

[1] Department of Informatics,
University of Beira Interior,
and IT - Instituto de Telecomunicações,
Covilhã, Portugal.

[2] ICS-UBI - Health Sciences Research Centre,
University of Beira Interior,
Covilhã, Portugal.

[3] Department of Chemistry,
University of Beira Interior,
Covilhã, Portugal.

Abstract

The current study presents an *in silico* model able to simulate the enzyme kinetics of a particular enzyme and substrate. For a given concentration of enzyme and substrate, the simulator calculates how quickly the enzyme stays saturated with a particular substrate, and the maximum rate it can achieve. These properties are important in order to infer the enzyme activity in the cell and to predict the way the enzyme will respond to changes in substrate concentration. Related studies can be found in literature most of them using complex continuous models such as partial differential equations. The simulator is based on a discrete model, easily customizable, which displays an animated representation of the enzyme-catalyzed reaction. In addition, through the double reciprocal representation from Lineweaver-Burk the parameters of Michaelis-Menten kinetics were obtained. Two types of inhibitors (irreversible and reversible competitive) were also added to the simulator.

1 Introduction

According to [4], there exists a wide range of possible problems to be modeled in Systems Biology. The chosen biological simulation problem was inspired by [2, 5] which models the enzyme kinetics represented by conventional biochemistry models namely, Michaelis-Menten [6] and Lineweaver-Burk [3]. The model used to study this particular biological problem was a discrete approach, the 2-D Cellular Automata (CA) with the extended Von Neumann neighborhood. The intent of the study is to analyze the capacity of Cellular Automata (CA) discrete models to simulate enzyme kinetics, using various amount of substrate concentration as input and study its influence in the results. However, this study was not only focused in the recognition of the ability to model enzyme kinetics, but also to adapt the model to an *in vitro* study [1] conducted at the chemistry laboratories at UBI. The simulator is based on a discrete model, displaying an animated representation of the enzyme-catalyzed reaction. Through the double reciprocal representation from Lineweaver-Burk the parameters of Michaelis-Menten kinetics were obtained. Two types of inhibitors (irreversible and reversible competitive) were also added to the simulator.

2 Methods

The methodology used for the development of the present work follows a step-by-step approach, and can be divided into four main parts: (1) understanding the potential of the CA models in systems biology resourcing to the work presented in literature; (2) study and implementation of a specific biological problem to build a conceptual model of the problem; (3) tuning the model for a specific *in vitro* study; (4) validation of the model implemented.

To proceed to the implementation of the *in silico* model we choose Processing and MATLAB programming language, and CA discrete models. The CA were originally proposed by John Von Neumann as an abstract simulation of self-reproduction in biology. These models run under a finite dimensional grid with three main concepts: (1) cells, that represent the entities; (2) neighbors, that introduce the interaction between entities; (3) rules, that bring the dynamics to the entire system. With the initial research in literature about the existing applications of CA in systems biology was possible to conclude that these models are a good approach and widely used in this field. We found a wide range of modeling applications of CA to systems biology such as in HIV, angiogenesis, cancer cells (in the study of tumor growth), metabolic pathways and many others. There are also several studies related to enzyme kinetics. Notwithstanding these studies served as the basis to this simulator, they are quite limited because did not use the inhibition component. The computer simulation presented here, operates in a 2D grid with a variable number of cells (usually used 12100 cells), representing enzymes, substrates, products and inhibitors (only if needed). The type of neighborhood chosen was the extended Von Neumann. As the created system also intended to implement the ability to modeling of different enzymes there was introduced a parameter that determine the probability to occur the catalysis (designed as Pc). In attempt to get a better understanding of the performing of the system figure 1 presents an example of the operations made during an iteration.

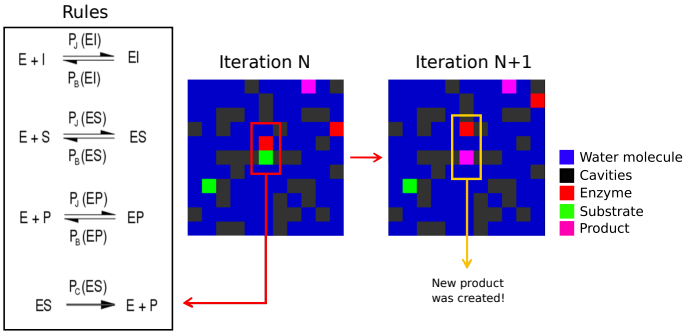


Figure 1: Example of system operation during a time step, evolving the respective application of the rules and the result in the next generation.

3 Results

In this section follows the results achieved with the current work. The output data is the result of three main target studies: the response of the system in the presence of different types of inhibitors (see fig. 2); the influence of each input parameter in the final result; and the ability of the system to portray an *in vitro* experiment.

3.1 *In vitro study*

The system suffured a set of tests in attempt to produce similar results to those achieved in [1]. As the system has a huge amount of different combinations of initial values, was previously made a set of trials in attempt to understand the influence of each parameter in the final outcomes. The most accurate results are shown below figures 2 and 3. This study can be viewed as a mathematical induction such as:

1. **Given statement:** CA implemented can approximate values from those obtained in *in vitro* experience;
2. **Base case:** test to a specific enzyme and substrate, Acetylcholinesterase and Acetylcholine respectively;
3. **Inductive step:** adjust the model to simulate other combinations of enzymes and substrates (not applicable in this study).

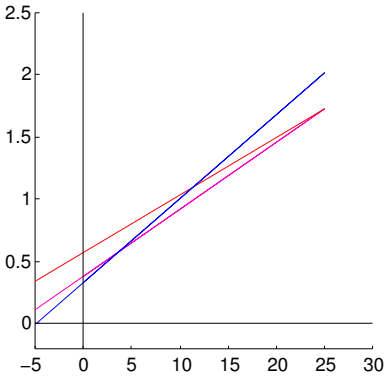


Figure 2: Representation of Lineweaver-Burk kinetics, where the pink line represents the "no inhibitors" case, the red line represents the "irreversible inhibitors" case and the blue line represents the "reversible competitive inhibitor".

The Table 1 shows the substrate and enzyme concentration used for this study. The total number of cells was 12100. The number of substrate concentrations were equal from those used in the *in vitro* experiment, however the enzyme concentration were induced by us, because was not mentioned in the reference document.

$[E]$	$[S]_1$	$[S]_2$	$[S]_3$	$[S]_4$	$[S]_5$	$[S]_6$	$[S]_7$
0.000347	0.0025	0.0035	0.005	0.0075	0.01	0.015	0.02

Table 1: Substrate concentration used in *in vitro* study.

The Table 2 presents the values achieved in this study. As it is possible to check exists a big difference between values. The most closer value obtained was the K_m, but the others are very distant from the desired. However, is important to note that Lineweaver-Burk plot is obtained using the reciprocal of v_0 values, so when the value of v_0 comes very small its reciprocal tends to be very large. The reader can get a graphical overview of the differences obtained in figure 3.

	In vitro	*In silico*	ERROR
K_m(mM)	0.0656	0.1973	0.1317
V_{max}(mM^{-1}.s)	0.0015	0.0047	0.0032

Table 2: Comparison values between *in vitro* results and the closer values obtained with CA model.

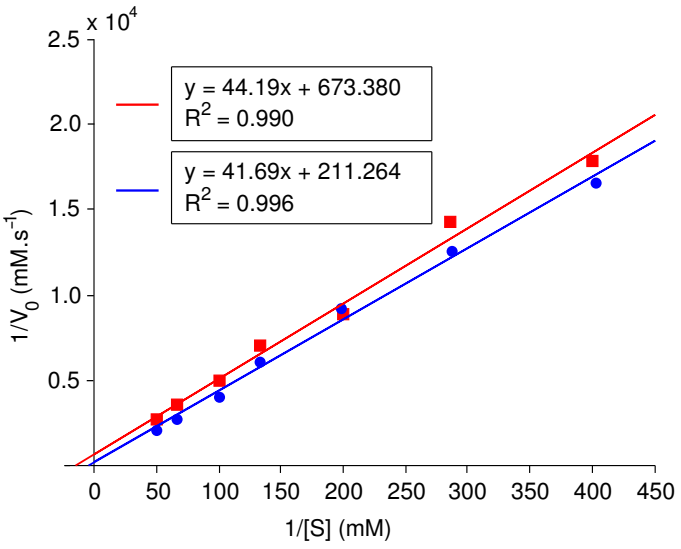


Figure 3: Representation of Lineweaver-Burk kinetics of the *in silico* model comparing to the *in vitro experiment*, where the blue line represents the value obtained in *in vitro* study and the red line the values obtained with *in silico* model.

The system was able to identify the same general overall trends as the ones identified from the available in vitro experiments. However, and as already mentioned, still exists an error in this model, which probably could be minimized with the introduction of more aspects to take into account such as the temperature. Other important aspect that could reduce this difference is a relation between unities such the number of interactions at each generation used in the *in silico* model and the time, in seconds, used in the *in vitro* experiment.

4 Conclusions

The CA models have the ability to reproduce, with an high level of realism, this specific biological problem - the enzymatic kinetics. From the literature review the wide applicational range of CA to modeling and simulation in Systems Biology is clear. The developed model was generated from appropriate features and reflects some interesting properties of enzyme kinetics, such as the influence of different input parameters in the final results.

However it is important to mention that the reported results refer to a specific pair of enzyme-substrate in a given time interval. When facing a more complex biological system such as the simulation of an entire metabolic pathway, which involve many different enzymes and substrates, the exclusive use of these discrete models should not suffice.

Nevertheless, the current work encourages further experiments and study, providing an useful introductory hands-on approach to enzyme kinetics using CA. As a complement to this research we intend to perform a parametric analysis of the influence of additional parameters such as the solvent, the temperature and the pH of the environment. The adjustment of the parameters to different *in vitro* studies appears to be another interesting line of work. This will result not only in a more realistic model of enzyme kinetics but also in a wider range of new applications.

References

[1] Nelson António Freitas Fernandes. *Aplicação da atividade biológica de compostos naturais de plantas medicinais de Angola e de compostos de síntese*. PhD thesis, University of Beira Interior, 2014.

[2] Lemont B. Kier, C.-K. Cheng, Bernard Testa, and Pierre-Alain Carrupt. A cellular automata model of enzyme kinetics. *Journal of Molecular Graphics*, 14(4):227 – 231, 1996. ISSN 0263-7855. doi: http://dx.doi.org/10.1016/S0263-7855(96)00073-2. URL `http://www.sciencedirect.com/science/article/pii/S0263785596000732`.

[3] Hans Lineweaver and Dean Burk. The determination of enzyme dissociation constants. *Journal of the American Chemical Society*, 56 (3):658–666, 1934.

[4] João Silva, Francisco Ribeiro, and Paulo Fazendeiro. Cellular Automata for Systems Biology: a Review, SOCIA Tech Report N. 01/2014. Technical report, University of Beira Interior, 2014.

[5] Jörg R. Weimar. Cellular Automata Approaches to Enzymatic Reaction Networks. In Stefania Bandini, Bastien Chopard, and Marco Tomassini, editors, *Cellular Automata*, volume 2493 of *Lecture Notes in Computer Science*, pages 294–303. Springer Berlin Heidelberg, 2002. ISBN 978-3-540-44304-9. doi: 10.1007/3-540-45830-1_28. URL `http://dx.doi.org/10.1007/3-540-45830-1_28`.

[6] H. W. Wiley. LOIS GÉNÉRALES DE L'ACTION DES DIASTASES. *Journal of the American Chemical Society*, 25(7):780–782, 1903. doi: 10.1021/ja02009a024.

Automatic Segmentation of Corneal Nerves from In Vivo Confocal Microscopy images using Chan-Vese Model and Tortuosity Evaluation

David Saraiva[1], Hugo Marcos[2]
{a21210488, a21210492} @alunos.isec.pt

Verónica Vasconcelos[3], António Luís Marques[3,4]
{veronica, lmarques} @isec.pt

[1,2] Biomedical Engineering Students
Coimbra Institute of Engineering, Polytechnic Institute of Coimbra

[3] Coimbra Institute of Engineering, Polytechnic Institute of Coimbra

[4] Institute of Telecommunications

Abstract

The incidence of Diabetes Mellitus (DM) is increasing dramatically in developed countries. One of the most common side effects of this disease is diabetic neuropathy which is the main responsible for the disability associated with Diabetes.

Using corneal confocal microscopy (CCM) is possible to obtain image with *in vivo* the sub-basal nerve plexus. Several studies show the importance of corneal nerves analysis in the early diagnostic of peripheral neuropathy, a frequent and severe side effect of Diabetes. In an early stage of the disease, is possible to detect a significant decrease in density of nerve and variations of morphometric parameters of the sub-basal nerves plexus, like nerve tortuosity, in CCM images of patients with diabetic neuropathy [1].

The development of automated systems that allow segmentation of the corneal nerves and calculus of specific morphometric parameters, like tortuosity, can be a simple and non-invasive way to early detect diabetic neuropathy.

In this paper is present an automatic segmentation of corneal nerves from CCM images using a new model for active contours based on curve evolution. After corneal nerve segmentation and post-processing techniques the tortuosity of nerves is computed.

1 Introduction

Diabetic neuropathy is a nerve damage disease that occurs in the body due to high levels of sugar in the blood. About half of all the patients with DM of type I and type II will develop neuropathy in some part of their body [2]. So an early diagnosis and accurate assessment of this pathology, using non-invasive methods, is of great importance for mundial health.

In this paper we propose an automatic segmentation algorithm for nerves of the cornea from images obtained by CCM and subsequent calculation of tortuosity of the segmented nerves. The evaluation of nerve tortuosity is related to neuropathy severity [3], can be useful on staging of diabetic peripheral neuropathy.

The Fig. 1 shows images of non-tortuous and tortuous corneal nerves obtained by CCM. The nerves appear bright in a dark background, frequently; these images have other types of structures like basal epithelial cells and keratocytes. The sub-basal nerve plexus images acquired with CCM have a central region brighter than the peripheral regions which difficult segmentation stage more difficult. Before nerve segmentation is essential to perform a correct pre-processing in order to enhance contrast and luminosity.

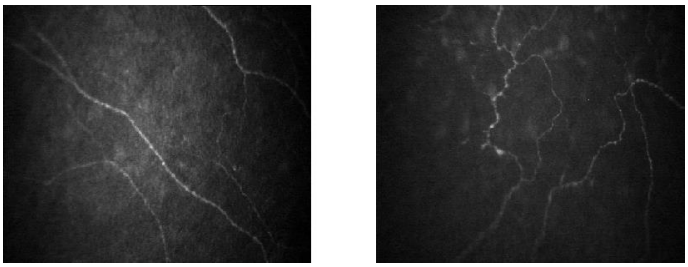


Figure 1: CCM Images of non-tortuous corneal nerves of a healthy subject (left) and tortuous sub-basal corneal nerves of a subject with diabetes (right.)

2 Methods

The proposed algorithm for automatic segmentation is divided in tree major stages: pre-processing, segmentation and post-processing.

2.1 Pre-Processing

The pre-processing stage was a challenge. Several methods were tested with different parameters. The best results were obtained by successive application of image processing methods. We started by apply the gray scale morphological Top-Hat transformation to correct uneven luminosity of the dark background. The top-hat transformation of a gray-scale image f by the structuring element b is defined as the original image f minus its opening that corresponds to two successive morphological operation: erosion $\otimes$ and dilation [4]:

$$T_{hat}(f) = f - (f \odot b) = f - (f \otimes b) \oplus b \quad (1)$$

where $\oplus$ and $\otimes$ represents the dilation and erosion, respectively.

After intense experimentation the chosen structuring element was a disk with radius of 13 pixels. To blur the image and attenuate the noise presence it was implemented a Gaussian lowpass filter of size 6x6 and standard deviation of 1.8. To blur the image and attenuate the noise presence it was implemented a Gaussian lowpass filter of size 6x6 and standard deviation of 1.8. To enhance the contrast the filtered images are transformed using contrast-limited adaptive histogram equalization (CLAHE). CLAHE operates on small tiles of the image rather than on the entire image. Square regions of 8 x 8 pixels were used.

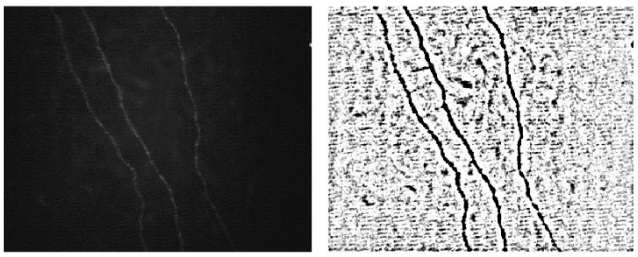


Figure 2: Original CCM image (left) and results of the pre-processing stage (right.

2.2 Segmentation

The process of image segmentation consists in partitioning an image into different regions based on specified criterion. Regions of interest are isolated in order to be represented and described by high level features simplifying the image processing in the specific problem context [4][5].

In a first approach segmentation was performed through adaptive thresholds using the Otsu's method. This method proved effective for some images but showed a poor ability to generalize. As the geometric and topological shapes of corneal nerves are not a priori known the choice Chan-Vese model [6] also known as Active Contours without Edges emerged as the right choice. This algorithm has been used successfully in the segmentation of different structures in medical images, especially for the segmentation of brain, heart and trachea [8].

Chan & Vese proposed a new model for active contours based on techniques of curve evolution to detect objects in an image, Mumford–Shah functional [8] for segmentation and level sets.

The Chan-Vese approach was applied to all available CCM images good results have been obtained. The Figure 3 (left) shows an example of corneal nerves segmentation using Chan-Vese method. The corresponding matrices to obtained images has positive and negative numbers, each one representing one level set of the image. The Figure 3 (right) is the result of binarization of the image obtained using the Chan-Vese method (left picture) limiting the value of the level set. In this case, values lower than 0.1 becomes black, while higher values becomes white.

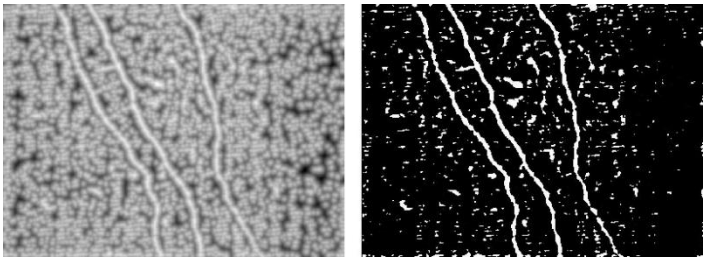


Figure 3: Segmentation using Chan-Vese model (left) and binary result of processing Chan-Vese segmentation results (right).

The obtained results can be explained by the fact that Chan-Vese approach is based on global properties rather than just taking into account local properties, such as gradients, which happens with the most methods of segmentation using level set methods.

2.3 Post-Processing

As expected, after the segmentation arise many parasites regions of small areas. So, a post-processing phase that isolates the region of interest - the nerves of the cornea - is required.

The remove of the unwanted areas in the binary images was performed in two steps. First the morphological operation close was applied to correct discontinuities of the nerves contributing to a better definition of the object under study because small details are captured and connected to the region of interest. This operation is composed by two simple morphological operations, dilation and erosion [5]:

$$f_{close} = f \bullet b = (f \oplus b) \otimes b \quad (2)$$

Then all segmented regions are labelled and their properties measured. Regions are removed based on area and eccentricity properties [9]. Using this strategy, almost one hundred percent of the parasite regions are removed emerging isolates nerves.

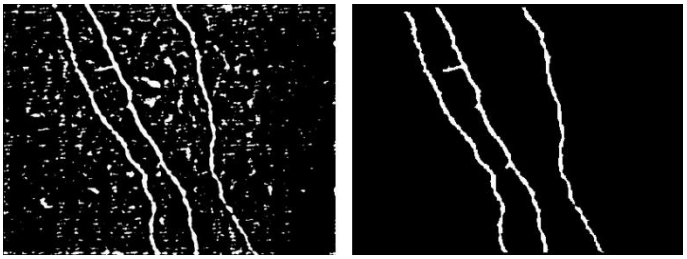


Figure 4: The binary image (left) and post-processed images using morphological operations (right).

Figure 4 shows an example of a binary image before (left) and after post-processing stage (right). The experiments showed the importance of the order of application of operations. First must be applied closing operation before deleting unwanted regions. In another way small areas belonging to the nerves would also be eliminated, reducing the continuity and detail of the segmented regions. In some situations the post-processing step can be refined, however for the purposes of calculating the nerve tortuosity these pixels parasites have no influence on the outcome.

2.4 Tortuosity

Tortuosity is a property usually associated with corneal nerves. It is a characteristic of the corneal nerve architecture that says how curved or kinked a nerve appears along its course. In diabetic patients the Tortuosity is a measure from neuropathy diabetic, so of great importance in the diagnosis and follow up of these patients.

There are several ways to determine the nerves tortuosity depending on the objective of the application.

In this study we choose a simple approach for the calculus of the tortuosity, because it is easy to understand and requires less computing cost, making the process faster. The considered method has the following expression:

$$\tau = \frac{L}{C} \quad (3)$$

Where L is the length of the considered nerve and C represents the distance between the ends of the nerve. The Figure 5 show a graphic interpretation of the tortuosity expression presented above.

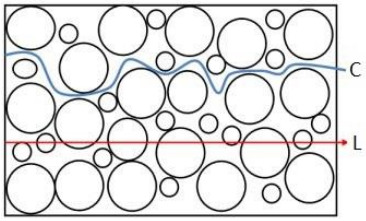


Figure 5: Length of the vessel (C) and distance between edges (L).

To implement this approach to calculate the corneal nerve tortuosity we had to make some modifications in the segmented image to achieve the pretended results. When the nerves have no ramifications the procedure is pretty straightforward. The problem occurs when the nerves have branches. The solution was to locate the branching points of branch nerves are isolated from each division in order to implement the algorithm presented above, as shown in Figure 6:

Figure 6: Segmented and processed image (left) and branch selected for tortuosity calculation (right).

3 Discussion

Figure 7 summarizes the application of the proposed methodology for segmentation of corneal nerves and calculation of its tortuosity. The right image shows an isolated nerve and the value of its tortuosity. The results are consistent with the visual analysis of the shape of the corneal nerves.

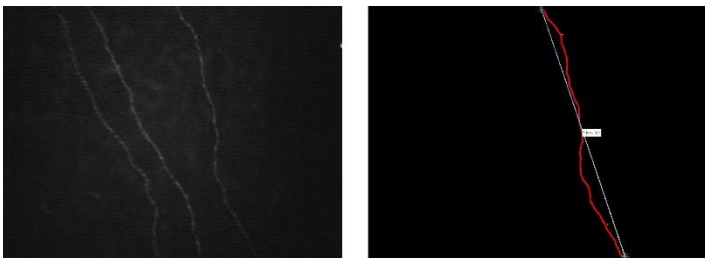


Figure 7: Original image (left) and final image with the segmented nerve and value of tortuosity (right).

Acknowledgment

The authors thank Dr. José Silva and Dr. Miguel Morgado for providing the images used in this study.

References

[1] Popper, M., Quadrado, M.J., Morgado, A.M., Murta, J.N., Van Best, J.A., Muller, L.J. Subbasal Nerves and Highly Reflective Cells in Corneas of Diabetic Patients: In vivo Evaluation by Confocal Microscopy. Invest. Ophthalmol. Vis. Sci., vol. 46, 2005.

[2] D. Quan. Diabetic Neuropathy. May, 2014. [Online]. Available: http://emedicine.medscape.com .

[3] K. Panagiotis, B. Michael, O'Donnell Clare, B. Andrew J. M., E. Nathan et al. *Corneal Nerve Tortuosity in Diabetic Patients with Neuropathy*. Invest. Ophthalmol. Vis. Sci., (45): 418-422, 2004.

[4] L. G. Shapiro and G. C. Stockman. *Computer Vision*. New Jersey, Prentice-Hall, 2001.

[5] R. C. Gonzalez and R. E. Woods. *Digital image processing*. Upper Saddle River, NJ : Pearson/Prentice Hall,2008.

[6] T. F. Chan, L. A. Vese. Active contour without edges. *IEEE Trans. On Image Processing*, 10(2):266-277, 2001.

[7] D. Mumford and J. Shah. Optimal approximation by piecewise smooth functions and associated variational problems. Comm. Pure Appl. Math, (42):577-685, 1989.

[8] O. Rousseau and Y. Bourgault. *Heart segmentation with an iterative Chan-Vese algorithm.* University of Ottawa, Ontario, 2009.

[9] O. Marques. *Practical image and video processing using MATLAB*, New Jersey, John Wiley & Sons, Inc., 2011.

Single-Frame Image Denoising and Inpainting using Gaussian Mixture Models and Minimum Mean Squared Error Estimation

Afonso Teodoro[1,3]
afonso.teodoro@tecnico.ulisboa.pt

Mariana Almeida[2,3]
mariana.almeida@lx.it.pt

Mário Figueiredo[1,3]
mario.figueiredo@tecnico.ulisboa.pt

(1) Instituto Superior Técnico, Universidade de Lisboa

(2) Priberam Labs, Lisboa

(3) Instituto de Telecomunicações, Lisboa

Abstract

This paper proposes a method that is able to address two of the major problems in image processing: denoising and inpainting. The approach is based on a Gaussian mixture model (GMM), estimated exclusively from the observed image, using the expectation-maximization algorithm, based on which the minimum mean squared error (MMSE) estimate is obtained in closed form. To the best of our knowledge, this approach has never been documented and the results show that it is able to perform on the same level as other state of the art algorithms.

1 Introduction

With the current technological development, there is an increasingly higher number of applications that, at some point, involve acquiring one or more images, be it for recreational, medical, surveillance, astronomical, or any one of many other purposes. Naturally, the development of methods to improve the quality of these images is a necessity.

Image denoising is a well-known problem of image processing; it was one of the first to be addressed and is still one of the core problems. As the name suggests, image denoising deals with the removal of noise from digital images. Another example refers to filling-in areas that are missing. The latter is commonly called image inpainting and, along with image denoising, is associated with the process of acquiring digital images.

1.1 Problem statement

The two aforementioned problems belong to a more general family of ill-posed problems, called inverse problems: there is some observed data and the objective is to determine a set of parameters which originated it, according to a theoretical model. Equation (1) is a typical formulation of such problems:

$$\mathbf{y} = H\mathbf{x} + \mathbf{n}, \tag{1}$$

where $\mathbf{y}$ denotes the corrupted input signal, $\mathbf{x}$ the clean image that we want to estimate, $\mathbf{n}$ is noise, and H is a degradation operator that depends on the type of problem at hand. In image denoising H is an identity matrix while in image inpainting it represents a mask. There are numerous approaches to estimate the clean image but this paper will focus only on a particular class of algorithms: patch-based. In image denoising, some of the best approaches try to find a sparse representation of the patches, either in a transform domain [3], such as wavelet or DCT, or using a dictionary [1], while others simply search the image for similar patches and combine them [2]. The proposed method is based on a probabilistic model that uses a mixture of Gaussian distributions to represent the (usually overlapping) image patches. Although Gaussian mixture models have been used before to model image patches, estimation is usually based on a collection of clean images, rather than only on the observed image to be denoised, which is the approach pursued in this work.

In Section 2, the algorithm is introduced and explained in detail. Then, in Section 3, the results are presented, discussed and compared with other state of the art algorithms. Lastly, in Section 4, some final remarks and considerations are drawn.

2 Method

This section describes the proposed method and the reasons why, hopefully, the output results will be in the same ballpark as the current state of the art. As mentioned above, the proposed algorithm tries to find a probabilistic model for the noise-free patches using mixture models. Such tool has proven to be very useful and flexible in modelling multivariate data. In particular, Gaussian mixture models (GMM) are capable of representing any continuous probability density function and have interesting analytical tractability properties [4]. Mathematically, a GMM is given by the well known expression

$$p_X(\mathbf{x} \mid \phi) = \sum_{i=1}^{k} \alpha_i \mathcal{N}(\mathbf{x}; \mu_i, C_i), \tag{2}$$

where $\mathcal{N}(\mathbf{x};\cdot)$ denotes the Gaussian (or normal) density distribution computed at $\mathbf{x}$, whereas μ_i and C_i are the mean vector and covariance matrix of the $i-th$ component. The parameters α represent the weight of each component, thus are non-negative quantities, summing to one. All the parameters that define the GMM (the set of all μ_i, C_i and α) are encompassed in ϕ. Estimating each clean image patch, $\mathbf{x}$, from the corresponding observed (noisy) patch $\mathbf{y}$, can then be cast as the following problems: first, estimate the parameters of the GMM from the set of observed image patches $\mathbf{y}_1, ..., \mathbf{y}_N$; second, compute the minimum mean squared error (MMSE) estimate of each patch $\mathbf{x}$, which is given by the posterior expected value:

$$\hat{\mathbf{x}} = E[\mathbf{x} \mid \mathbf{y}] = \int \mathbf{x} \frac{p_{Y|X}(\mathbf{y} \mid \mathbf{x})\, p_X(\mathbf{x})}{p_Y(\mathbf{y})} d\mathbf{x}. \tag{3}$$

The first problem can be solved by determining the maximum likelihood (or maximum *a posteriori*, in the presence of some prior) estimates of the mixture parameters. Although this can not be done analytically, it can be performed efficiently using the widely acknowledged *expectation-maximization* (EM) algorithm [4]. As for the second problem, Equation 3 is obtained by application of the Bayes rule and by minimizing the posterior squared error [7]. In general, the MMSE estimate is harder to compute, when compared to the maximum *a posteriori* estimate, due to the need to compute $p_Y(\mathbf{y})$. Yet, if the noise is additive white Gaussian noise, with zero mean and covariance matrix $\sigma^2 I$ ($\mathbf{n} \sim \mathcal{N}(0, \sigma^2 I)$), then the calculations greatly simplify. Considering the image denoising theoretical model (Equation (1)) and the probability distributions of $\mathbf{x}$ and $\mathbf{n}$, the following holds:

$$p_X(\mathbf{x}) = \sum_i \alpha_i \mathcal{N}(\mathbf{x}; \mu_i, C_i) = \sum_i \alpha_i p_X^i(\mathbf{x}), \tag{4}$$

$$p_Y(\mathbf{y}) = \sum_i \alpha_i \mathcal{N}(\mathbf{y}; \mu_i, C_i + \sigma^2 I) = \sum_i \alpha_i p_Y^i(\mathbf{y}), \tag{5}$$

$$p_{Y|X}(\mathbf{y} \mid \mathbf{x}) = \mathcal{N}(\mathbf{y}; \mathbf{x}, \sigma^2 I). \tag{6}$$

Substituting Equations (4), (5) and (6) in (3), a somewhat lengthy process of algebraic manipulation leads to

$$\hat{\mathbf{x}} = \sum_{i=1}^{k} \alpha_i'(\mathbf{y})\, \mu_i'(\mathbf{y}), \tag{7}$$

where

$$\alpha_i'(\mathbf{y}) = \frac{\alpha_i\, p_Y^i(\mathbf{y})}{p_Y(\mathbf{y})}, \tag{8}$$

are, once again, non-negative and sum to one, and

$$\mu_i'(\mathbf{y}) = C_i' \left(C_i^{-1} \mu_i + (\sigma^2 I)^{-1} \mathbf{y} \right), \tag{9}$$

(a) (b) (c) (d)

Figure 1: Denoising: (a) original image; (b) noisy image ($\sigma = 30$, PSNR = 18.59dB); (c) BM3D (PSNR = 31.26dB); (d) Proposed (PSNR = 30.99).

and

$$C_i^{'} = \left(C_i^{-1} + (\sigma^2 I)^{-1}\right)^{-1}, \tag{10}$$

are obtained knowing that the product of two Gaussian distributions is proportional to another Gaussian distribution with mean and covariance matrix given by Equations (9) and (10) (see, for example, [5]). Note that, as expected, both $\alpha_i^{'}$ and $\mu_i^{'}$ depend on the input data.

All the steps above are also applicable to the image inpainting problem and the final result is the same as Equation (7), with a slight modification to account for the missing data. Matrix H is, in this situation, a binary mask with 1's in the observed pixels and 0's elsewhere. An alternative way to write Equation (1), in the case of inpainting is

$$\mathbf{y}_{obs} = M\mathbf{x} + \mathbf{n}, \tag{11}$$

where $\mathbf{y}_{obs}$ corresponds to the observed pixels only and M is a fat matrix, also binary, such that $MM^T = I$ and $M^T M = H$. These last two properties of matrix M are necessary to conclude that

$$\mu_i^{'}(\mathbf{y}_{obs}) = C_i^{'}\left(C_i^{-1}\mu_i + M^T(\sigma^2 I)^{-1}\mathbf{y}_{obs}\right), \tag{12}$$

where

$$C_i^{'} = \left(C_i^{-1} + M^T(\sigma^2 I)^{-1}M\right)^{-1}. \tag{13}$$

3 Results and discussion

In this Section, experimental results obtained with the proposed method are presented and discussed. Tables 1 and 2 summarize the numerical results of image denoising and inpainting, while Figure 1 shows an example of the visual quality of the output. Table 1 was obtained through an extensive analysis of the impact of the algorithm parameters, namely the size of the patches ($pd \times pd$) and the number of components in the mixture (K). The experiments were carried out as follows: for each input image and noise variance, both pd and K were swept from 3 to 12 and 10 to 60 (intervals of 5), respectively. Such experiment allowed us to draw some conclusions: the values of the parameters that provided best results depend on the noise level, on the size of the image, and on whether the image is flat or rich in texture. These dependencies are a downside of the proposed method and a strategy to make the algorithm more robust should be developed. Even so, the algorithm described in Section 2 performs on the same level as other state of the art algorithms.

The inpainting experiment (Table 2) was simpler taking into account the previous considerations: all the values were obtained with fixed patch size ($pd = 8$) and number of components ($K = 25$). As expected, as the data ratio decreases the harder it gets to recover the missing pixels. Still, the proposed method produced state of the art results.

Note: the four methods used for comparison were chosen for two reasons: first, they provide state of the art results on each problem, and second, there is a publicly available implementation of the algorithms.

4 Conclusions

From what could be ascertained, this is the first time that a Gaussan mixture model obtained with the EM algorithm exclusively from the observed image (rather than from a training set of clean images) is used to address problems such as image denoising and inpainting. The obtained results are promising but we believe that there is still room for improvement. Nonetheless, this paper proposed an alternative approach, which was shown to be able to provide state of the art results.

sigma (σ)	Lena (512 × 512)			Cameraman (256 × 256)			House (256 × 256)		
	BM3D	K-SVD	Proposed	BM3D	K-SVD	Proposed	BM3D	K-SVD	Proposed
5	38.72	38.53	**38.86**	38.29	37.97	**38.57**	39.83	39.47	**39.92**
10	**35.93**	35.55	35.88	34.18	33.76	**34.44**	**36.61**	36.05	36.58
15	**34.27**	33.74	34.11	31.91	31.54	**32.15**	**34.94**	34.41	34.65
20	**33.05**	32.40	32.83	30.48	30.07	**30.64**	**33.77**	33.21	33.34
25	**32.08**	31.34	31.81	29.45	28.94	**29.50**	**32.86**	32.21	32.34
30	**31.26**	30.46	30.99	**28.64**	28.12	28.58	**32.09**	31.25	31.44

Table 1: PSNR comparison on gray-level image denoising.

Method	Lena (512 × 512)			Cameraman (256 × 256)			House (256 × 256)		
	80%	50%	20%	80%	50%	20%	80%	50%	20%
BP [8]	41.27	36.94	31.00	34.70	28.90	24.11	43.03	38.02	30.12
SO [6]	**43.01**	**37.43**	**31.94**	**36.19**	**30.71**	**25.13**	44.34	38.77	**32.60**
Proposed	41.91	37.22	31.25	33.74	28.85	24.11	**45.20**	**39.30**	31.71

Table 2: PSNR comparison on gray-level image inpainting.

Acknowledgments

This work was partially supported by *Fundação para a Ciência e Tecnologia* (FCT), under grants PEst-OE/EEI/LA0008/2013 and PTDC/EEIPRO/1470/2012.

References

[1] M. Aharon, M. Elad, and A. Bruckstein. K-SVD: An Algorithm for Designing Overcomplete Dictionaries for Sparse Representation. *Signal Processing, IEEE Transactions on*, 54(11):4311–4322, 2006.

[2] A. Buades, B. Coll, and J. M. Morel. Non-Local Means Denoising. *Image Processing On Line*, 2011, 2011.

[3] K. Dabov, A. Foi, V. Katkovnik, and K. Egiazarian. Image denoising by sparse 3D transform-domain collaborative filtering. *IEEE Transactions on image processing*, 16(8), 2007.

[4] M. Figueiredo and A. K. Jain. Unsupervised learning of finite mixture models. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 24:381–396, 2002.

[5] K. B. Petersen and M. S. Pedersen. *The Matrix Cookbook*. 2012.

[6] I. Ram, M. Elad, and I. Cohen. Image processing using smooth ordering of its patches. *CoRR*, abs/1210.3832, 2012.

[7] H. Van Trees. *Detection, Estimation, and Modulation Theory (Part I)*. Wiley, 1968.

[8] M. Zhou, H. Chen, J. Paisley, L. Ren, L. Li, Z. Xing, D. Dunson, G. Sapiro, and L. Carin. Nonparametric Bayesian dictionary learning for analysis of noisy and incomplete images. *Image Processing, IEEE Transactions on*, 21(1):130–144, 2012.

A 3D Keypoint Detector based on Biologically Motivated Bottom-Up Saliency Map

Sílvio Filipe
http://socia-lab.di.ubi.pt/~silvio
luís A. Alexandre
http://di.ubi.pt/~lfbaa

IT - Instituto de Telecomunicações
Department of Computer Science
University of Beira Interior
6200-001 Covilhã, Portugal

Abstract

We present a new method for the detection of 3D keypoints on point clouds and we perform benchmarking between each pair of 3D keypoint detector and 3D descriptor to evaluate their performance on object and category recognition. Our keypoint detector is inspired by the behavior and neural architecture of the primate visual system. The 3D keypoints are extracted based on a bottom-up 3D saliency map, that is, a map that encodes the saliency of objects in the visual environment. The saliency map is determined by computing conspicuity maps of the orientation, intensity and color information in a bottom-up and in a purely stimulus-driven manner. Finally, the focus of attention (or "keypoint location") is sequentially directed to the most salient points in this map. The main conclusions are: with a similar average number of keypoints, our 3D keypoint detector outperforms the other eight 3D keypoint detectors evaluated in the category and object recognition experiments.

1 Introduction

The interest on using depth information on computer vision applications has been growing recently due to the decreasing prices of 3D cameras. Depth information improves object perception, as it allows for the determination of its shape or geometry.

This paper has two main focuses: the first is to present a new keypoint detector; the second an evaluation of our and others 3D keypoint detectors. Our keypoint detector is a saliency model based on spatial attention derived from the biologically plausible architecture. It uses three feature channels: color, intensity and orientation. The computational algorithm of this saliency model has been presented in [6] and the standard saliency benchmark in 2D images. We present the 3D version of this saliency detector and demonstrate how keypoints can be extracted from a saliency map.

In [2], the author focuses on the descriptors available in Point Cloud Library (PCL), explaining how they work and made a comparative evaluation on publicly available data. In our study, we will see that it is not enough, the results also depend on the keypoint location. The same author studies the accuracy of the distances both for objects and category recognition and finds that simple distances give competitive results [1].

The 3D keypoint detectors and descriptors that we will compare can be found in version 1.7 of the PCL [9]. With this, we will find what is the best pair of keypoint detector/descriptor for 3D point cloud objects. We propose to answer this question using a public large RGB-D Object Dataset [7], this is composed of 300 real objects and divided in 51 categories.

In [4], the invariance of 3D keypoint detectors according to rotations, scale changes and translations was evaluated. It also contains a more detailed description of the two keypoint detectors: 1) The Scale Invariant Feature Transform (SIFT) keypoint detector was proposed by [8]. In [5], the original algorithm for 3D data is presented, which uses a 3D version of the Hessian to select the interest points; 2) Intrinsic Shape Signatures 3D (ISS3D) [13] is a method relying on region-wise quality measurements. This method uses the magnitude of the smallest eigenvalue (to include only points with large variations along each principal direction) and the ratio between two successive eigenvalues (to exclude points having similar spread along principal directions). We compare our proposal against these ones.

One of our goals was to evaluate the four descriptors, two main descriptors and its variants based on color, in terms of category and object recognition: 1) Descriptors such as Point Feature Histograms (PFH) [10] can be categorized as geometry-based descriptors. This type of descriptor is represented by the surface normals, curvature estimates and distances, between point pairs. PFHRGB is an version of PFH in which is included information regarding the color of the object; 2) The Signature of Histograms of OrienTations (SHOT) descriptor [11] is based on a signature histograms representing topological features, that make it invariant to translation and rotation. For a given keypoint, it computes a repeatable local reference frame using the eigenvalue decomposition around it. In order to incorporate geometric information of point locations in a spherical grid. For each spherical grid bin, a a one-dimensional histogram is obtained. In [12], they propose a color version (SHOTCOLOR), where use the CIELab color space as color information.

2 Proposed 3D Keypoint Detector

The Biologically Inspired 3D Keypoint based on Bottom-Up Saliency (BIK-BUS) is a keypoint detector that is based on the saliency maps, which are also known as visual attention [3]. The saliency maps are determined by computing conspicuity maps of the features intensity and orientation in a bottom-up and data-driven manner. These conspicuity maps are fused into a saliency map and, finally, the focus of attention is sequentially directed to the most salient points in this map. Using this theory and following the work presented in [6], we will present our keypoint detector in six steps.

Step 1: Linear Filtering – The color channels (r, g, and b) of the input colored point cloud are normalized when $I = (r+g+b)/3$ is larger than $1/10$ of its maximum over the entire image, other locations yield zero. With these three normalized color channels, we create four broadly-tuned color channels: $R = r-(g+b)/2$, $G = g-(r+b)/2$, $B = b-(r+g)/2$ and $Y = (r+g)/2-|r-g|/(2-b)$.

Five Gaussian pyramids $R(\sigma)$, $G(\sigma)$, $B(\sigma)$, $Y(\sigma)$ and $I(\sigma)$ are created from the color and intensity channels, where σ represents the standard deviation used in the Gaussian kernel. Each Gaussian pyramid is achieved by convolving the cloud with Gaussian kernels of increasing radius, resulting in a pyramid of clouds.

The orientation pyramids $O(\sigma,\theta)$ are obtained using the normals extracted from the intensity cloud I, where $\theta \in \{0^o, 45^o, 90^o, 135^o\}$ is the preferred orientation. In the primary visual cortex, the impulse response of orientation-selective neurons is approximated by Gabor filters.

Step 2: Center-Surround Differences – There are two types of center-surround structures in the retina: *on-center* and *off-center*. The *on-center* use a positive weighed center and negatively weighed neighbors and the *off-center* use exactly the opposite. The positive weighing is better known as excitatory and the negative as inhibitory [3].

Center-surround is implemented in the model as the difference between fine and coarse scales: the center is a pixel at scale $c \in \{2,3,4\}$, and the surround is the corresponding pixel at scale $s = c+\delta$, with $\delta \in \{3,4\}$. The across-scale difference between two maps, denoted by '$\ominus$', is obtained by interpolation to the finer scale and point-by-point subtraction.

The first set of feature maps is concerned with intensity contrast. Here, both types of sensitivities are simultaneously computed in a set of six maps $I(c,s) = |I(c) \ominus I(s)|$. For the color channels, the process is similar, which, in the cortex, is called 'color double-opponent' system. The existence of a spatial and chromatic opponency between color channels in human primary visual cortex is described. With that, the maps $RG(c,s)$ and $BY(c,s)$ are created in the model to simultaneously take in account the red/green and green/red, and blue/yellow and yellow/blue double opponency, respectively, as: $RG(c,s) = |(R(c)-G(c)) \ominus (G(s)-R(s))|$ and $BY(c,s) = |(B(c)-Y(c)) \ominus (Y(s)-B(s))|$. Orientation feature maps, $O(c,s,\theta)$, encode local orientation contrast between the center and surround scales: $O(c,s,\theta) = |O(c,\theta) \ominus O(s,\theta)|$.

Step 3: Normalization – The salient objects appear only in a few maps, which can be masked by noise or by less-salient objects present in a larger number of maps. In the absence of top-down supervision, we use a map normalization operator $\mathcal{N}(.)$ and consists of: 1 – Normalizing the values in the map to a fixed range $[0..M]$, in order to eliminate large amplitude differences; 2 – Finding the location of the global maximum maps M and computing the average $\overline{m}$ of all its other local maxima; and 3

– Globally multiply the map by $(M-\overline{m})^2$.

Step 4: Across-Scale Combination – The conspicuity maps are the combination of the feature maps, for intensity, color, and orientation, at the scale s = 4 of the saliency map. They are obtained through the reduction of each map to scale four and point-by-point addition, called across-scale addition, '$\bigoplus$'. The conspicuity maps for the intensity, $\overline{I}$, and color channels, $\overline{C}$, are given by: $\overline{I} = \bigoplus_{c=2}^{4} \bigoplus_{s=c+3}^{c+4} \mathcal{N}(I(c,s))$ and $\overline{C} = \bigoplus_{c=2}^{4} \bigoplus_{s=c+3}^{c+4} [\mathcal{N}(RG(c,s)) + \mathcal{N}(BY(c,s))]$.

For orientation, four intermediary maps are first created by combination of the six feature maps for a given θ and are then combined into a single orientation conspicuity map $\overline{O} = \sum_{\theta \in \{0^o, 45^o, 90^o, 135^o\}} \mathcal{N}\left[\bigoplus_{c=2}^{4} \bigoplus_{s=c+3}^{c+4} \mathcal{N}(O(c,s,\theta))\right]$.

Step 5: Linear Combination – The final saliency map is obtained by the normalization of each conspicuity map and calculating the mean between $\overline{I}$, $\overline{C}$ and $\overline{O}$.

Step 6: Inhibition-of-Return (IR) – The IR is part of the method that is responsible for the selection of keypoints. It detects the most salient location and directs attention towards it, considering that location a keypoint. After that, the IR mechanism transiently suppresses this location in the saliency map and its neighborhoods in a small radius, such that attention is autonomously directed to the next most salient image location. Computationally, the IR performs a similar process of selecting the global and local maximums.

3 Experimental Evaluation and Discussion

In order to perform this evaluation, we will use three measures, which are the Area Under the ROC Curve (AUC) and the decidability (DEC). The obtained AUC and DEC for category and object recognition are given in table 1. The decidability index represents the distance between the distributions obtained for the two classical types of comparisons: between descriptors extracted from the same (*intra-class*) and different objects (*inter-class*). Considering μ_{intra} and μ_{inter} denote the means of the intra- and inter-class comparisons, σ^2_{intra} and σ^2_{inter} the respective standard deviations and the decidability is given by $DEC = \frac{|\mu_{intra} - \mu_{inter}|}{\sqrt{\frac{1}{2}(\sigma^2_{intra} + \sigma^2_{inter})}}$.

Analyzing the descriptors in a generic way, the best results were obtained with the PFHRGB. It is interesting to compare it to the PFH: improvement can only be attributed to the incorporation of color information. The same is true for the SHOTCOLOR versus the SHOT descriptor. The two best results in terms of category and object recognition are presented in the descriptors that use color information.

Considering only the accuracy, the best combination for the category recognition is BIK-BUS/PFHRGB, closely followed by BIK-BUS/ SHOTCOLOR, ISS3D/PFHRGB and ISS3D/SHOTCOLOR both in terms of AUC and DEC. The pairs BIK-BUS/PFHRGB and BIK-BUS/ SHOTCOLOR have exactly the same AUC, the difference is in the DEC where it is slightly higher in the case of PFHRGB. BIK-BUS turns out again the best performer among detectors: SHOT, SHOTCOLOR and PFHRGB.

In terms of object recognition, the best pair is BIK-BUS/PFHRGB, but only beats the second best combination, ISS3D/PFH. For SHOT and SHOTCOLOR descriptors, if we compare our keypoint detector with the ISS3D we obtain improvements for both of 1.5% in the case of category recognition, and 1.1% and 1.4% in object recognition, respectively.

4 Conclusions

In this paper we presented a novel 3D keypoint detector biologically motivated by the behavior and the neuronal architecture of the early primate visual system. The BIK-BUS is a keypoint detector on a computational technique to determine visual attention, which are also known as saliency maps. The saliency maps are determined by sets of features in a bottom-up and data-driven manner. The fusion of these sets produced the saliency map and the focus of attention is sequentially directed to the most salient points in this map, representing a keypoint location.

In the evaluation, we used some of the 3D keypoint detectors and the 3D descriptors available in the PCL library. The main conclusions of this paper are: 1) a descriptor that uses color information should be used instead of a similar one that uses only shape information; 2) in terms of keypoint detectors, to obtain an accurate recognition system, we recommend the use of the BIK-BUS, since its performance is by far the best among the keypoint detectors tested; 3) in terms of descriptors, if the focus is on accuracy we recommend the use of PFHRGB and for real-time a good choice is the SHOTCOLOR because it presents a good balance between recognition performance and time complexity.

Table 1: AUC and DEC values for the category and object recognition for each pair keypoint detector/descriptor. **BOLD** indicates the best (bigger) results in terms of AUC and DEC for each pair.

Descriptor	Keypoint	Category		Object	
		AUC	DEC	AUC	DEC
SHOT	BIK-BUS	**0.827**	**1.281**	**0.863**	**1.513**
	ISS3D	0.812	1.168	0.852	1.413
	SIFT3D	0.814	1.207	0.848	1.409
SHOTCOLOR	BIK-BUS	**0.867**	**1.571**	**0.916**	**2.012**
	ISS3D	0.852	1.465	0.902	1.873
	SIFT3D	0.839	1.394	0.896	1.792
PFH	BIK-BUS	**0.848**	1.488	0.893	1.832
	ISS3D	**0.848**	**1.489**	**0.895**	**1.855**
	SIFT3D	0.843	1.458	0.890	1.801
PFHRGB	BIK-BUS	**0.867**	**1.586**	**0.948**	**2.397**
	ISS3D	0.866	1.585	**0.948**	2.394
	SIFT3D	0.861	1.546	0.946	2.373

Acknowledgment

This work is supported by '*FCT - Fundação para a Ciência e Tecnologia*' (Portugal) through the research grant 'SFRH/BD/72575/2010', and the funding from 'FEDER - QREN - Type 4.1 - *Formação Avançada*', subsidized by the European Social Fund and by Portuguese funds through 'MCTES'. We also acknowledge the support given by the *IT - Instituto de Telecomunicações* through 'PEst-OE/EEI/LA0008/2013'.

References

[1] L. A. Alexandre. Set Distance Functions for 3D Object Recognition. In *18th ICPR*, pages 57–64. Springer, 2013.

[2] Luís A. Alexandre. 3D descriptors for object and category recognition: a comparative evaluation. In *Workshop on Color-Depth Camera Fusion in Robotics at the IEEE/RSJ IROS*, Vilamoura, Portugal, October 2012.

[3] S. Filipe and L. A. Alexandre. From the human visual system to the computational models of visual attention: a survey. *AIRE*, January 2013 (accepted).

[4] S. Filipe and L. A. Alexandre. A Comparative Evaluation of 3D Keypoint Detectors in a RGB-D Object Dataset. In *9th VISAPP*, Lisbon, Portugal, 5–8 January 2014.

[5] A. Flint, A. Dick, and A. Hengel. Thrift: Local 3D Structure Recognition. In *9th DICTA*, pages 182–188, December 2007.

[6] L. Itti, C. Koch, and E. Niebur. A model of saliency-based visual attention for rapid scene analysis. *IEEE TPAMI*, 20(11):1254–1259, March 1998.

[7] K. Lai, L. Bo, X. Ren, and D. Fox. A large-scale hierarchical multi-view RGB-D object dataset. In *ICRA*, pages 1817–1824, May 2011.

[8] D. G. Lowe. Distinctive image features from scale-invariant keypoints. *IJCV*, 60(2):91–110, November 2004.

[9] R. B. Rusu and S. Cousins. 3D is here: Point Cloud Library (PCL). In *ICRA*, Shanghai, China, May 9-13 2011.

[10] R. B. Rusu, N. Blodow, Z. C. Marton, and M. Beetz. Aligning point cloud views using persistent feature histograms. In *IEEE/RSJ IROS*, pages 3384–3391, Nice, France, September 2008.

[11] F. Tombari, S. Salti, and L. Di Stefano. Unique Signatures of Histograms for Local Surface Description. In *11th ECCV*, pages 356–369, Crete, Greece, 2010.

[12] F. Tombari, S. Salti, and L. Di Stefano. A combined texture-shape descriptor for enhanced 3D feature matching. In *18th IEEE ICIP*, pages 809–812, Brussels, September 2011.

[13] Y. Zhong. Intrinsic shape signatures: A shape descriptor for 3D object recognition. *ICCV*, pages 689–696, September 2009.

Author Index